

Helsingin yliopisto  
Slavistiikan ja baltologian laitos  
Venäjän kääntäminen ja tulkkaus

## **Merenkulun turvallisuuden koetinkiviä**

*Terminologisen tiedon poiminta teksteistä*

Päivi Pasanen

Esitetään Helsingin yliopiston humanistisen  
tiedekunnan suostumuksella julkisesti tarkastettavaksi  
auditoriumissa XII perjantaina 9. lokakuuta 2009 klo 12.

# **Merenkulun turvallisuuden koetinkiviä**

*Terminologisen tiedon poiminta teksteistä*

Päivi Pasanen

Helsinki University Translation Studies

Monographs 5

Helsinki 2009

Helsinki University Translation Studies  
Monographs 5

Vol. 5 Päivi Pasanen

Merenkulun turvallisuuden koetinkiviä. Terminologisen tiedon poiminta teksteistä

Kannen kuva: Tapani Pasanen 2009

© Päivi Pasanen

ISSN 1459-3246

ISBN 978-952-10-5672-7 (nid.)

ISBN 978-952-10-5673-4 (PDF) <http://ethesis.helsinki.fi/>

Yliopistopaino

Helsinki 2009

*Saaralle,*  
joka ottaa ensi askeleitaan  
elämänmittaisella tutkimusmatkalla

## ABSTRACT

### **Knowledge probes for maritime safety**

#### *The extraction of terminological information from texts*

The methodology of extracting information from texts has widely been described in the current literature. However, the methodology has been developed mainly for the purposes of other fields than terminology science. In addition, the research has been English language oriented. Therefore, there are no satisfactory language-independent methods for extracting terminological information from texts.

The aim of the present study is to form the basis for a further improvement of methods for extraction of terminological information. A further aim is to determine differences in term extraction between subject groups with or without knowledge of the special field in question. The study is based on the theory of terminology, and has mainly a qualitative approach. The research material consists of electronically readable specialized texts in the subject domain of maritime safety. Textbooks, conference papers, research reports and articles from professional journals in Finnish and in Russian are included.

The thesis first deals with certain term extraction methods. These are manual term identification and semi-automatic term extraction, the latter of which was carried out by using three commercial computer programs. The results of term extraction were compared and the recall and precision of the methods were evaluated. The latter part of the study is dedicated to the identification of concept relations. Certain linguistic expressions, which some researchers call knowledge probes, were applied to identify concept relations.

The results of the present thesis suggest that special field knowledge is an advantage in manual term identification. However, in the candidate term lists the variation between subject groups was not as remarkable as it was between individual subjects. The term extraction software tested here produces candidate term lists which can be useful, but only after some manual work. Therefore, the work emphasizes the need to further develop term extraction software. Furthermore, the analyses indicate that there are a certain number of terms which were extracted by all the subjects and the software. These terms we call core terms. As the result of the experiment on linguistic expressions which signal concept relations, a proposal of Finnish and Russian knowledge probes in the field of maritime safety was made. The main finding was that it would be useful to combine the use of knowledge probes with semi-automatic term extraction since knowledge probes usually occur in the vicinity of terms.

## ALUKSI

Kiitän lämpimästi kaikkia, jotka eri tavoin ovat edesauttaneet tämän työn valmistumista. Erityisesti haluan kiittää työni ohjaajia professori Inkeri Vehmas-Lehtoa ja professori Aleksandr Gerdiä. He luotsasivat minua eteenpäin tutkimustyön karikkoisilla vesillä silloinkin, kun omat navigointitaidot eivät riittäneet. Työni esitarkastajia professori Merja Koskela ja professori Mirja Kalasniemeä kiitän heiltä saamastani arvokkaasta palautteesta ja perinpohjaisesta paneutumisesta työni käsikirjoitukseen, joka ei ollut sivumäärältään aivan pieni. Ajautumisesta terminologisen tutkimuksen pariin kiitän professori Lauri Carlsonia, jonka kurssilla kiinnostukseni terminologiaan heräsi.

Suuret kiitokset Kouvolan kaupungille ja tutkijakoulu Langnetille, joiden rahoituksen turvin sain työskennellä, ja Helsingin yliopiston käännöstieteen laitokselle, joka tarjosi erinomaisen työskentely-ympäristön. Ensimmäiset askeleeni tieteen poluille sain ottaa Kouvolan tutkimusryhmässä. Haluan kiittää kaikkia siinä mukana olleita neuvoista ja aiheellisesta kritiikistäkin. Kiitokseni saavat myös kirjastonhoitajat Leena Eteläpää ja Ulla Seppälä sekä laitoksen koko henkilökunta, jonka tuki on ollut korvaamaton apu työssäni.

Kiitokset myös Helsingin yliopiston tutkimusvararyhmälle, Merenkulun säätiölle, Haminan yhteislyseon senioreille ja Kymin Osakeyhtiön 100-vuotissäätiölle saamistani apurahoista. Samoin suurkiitokset NordLing-tutkijakoululle, jonka apurahan turvin sain työskennellä 2,5 kuukautta GradEast-tutkijakoulun vieraana Kööpenhaminan kauppa-korkeakoulussa keväällä 2007. Kiitän Bodil Nistrup Madsenia ja Hanne Erdmann Thomsenia ohjauskeskusteluista sekä kaikkia Institut for Internationale Sprogstudier og Vidensteknologi -osaston työntekijöitä erittäin antoisasta ja uusia ajatuksia herättäneestä ajasta Kööpenhaminassa.

Erityiskiitokset Aleksandr Ševjakoville arvokkaasta tuesta venäjänkielisen aineiston kokoamisessa, suomalaisille ja venäläisille asiantuntijoille, jotka antoivat arvokasta aikaansa termilistan kokoamiseen, sekä suomalaisille ja venäläisille opiskelijoille, jotka osallistuivat terminpoimintaan.

Kiitän ystäviäni ja kollegoitani, joilta olen saanut arvokasta tukea näiden vuosien aikana. Kääntäjä Paula Backmania Merenkululaitoksesta kiitän yhteistyöstä, joka alkoi jo maisteriopintojen aikana ja toivottavasti jatkuu edelleen monin tavoin. Erityisesti kiitän Hanna Niemelää, joka jäi tuekseni sillä hetkellä, kun eniten ystävää tarvitsin, Igor Kudashevia, joka on pitänyt huolen, ettei minulta koskaan ole puuttunut venäjänkielistä luettavaa, sekä Marjut Alhoa ja Heidi Rontua, joiden kannustus on antanut voimia jatkaa eteenpäin.

Lämpimin ajatuksin kiitän vanhempiani Raunia ja Tapiota, aviomiestäni Tapania, joka on kriittisinäkin hetkinä jaksanut valaa minuun uskoa, että työ valmistuu, lapsiani Perttiä ja Riittaa, jotka kävivät koulun loppuun, opiskelivat ammattiin ja Riitta jopa perusti perheen sillä välin, kun minä vain laskin termejä, ja muita sukulaisiani, jotka ovat pitäneet minut jotenkin kiinni normaalielämässä.

Haminassa 9.8.2009

# SISÄLLYSLUETTELO

<b>ABSTRACT .....</b>	<b>4</b>
<b>ALUKSI.....</b>	<b>5</b>
<b>LYHENTEET .....</b>	<b>9</b>
<b>KUVIOT JA TAULUKOT .....</b>	<b>10</b>
<b>1 JOHDANTO .....</b>	<b>13</b>
1.1 Terminologinen tutkimus ja sanastotyö tiedon hallinnassa.....	13
1.2 Työn tarkoitus .....	15
1.3 Merenkulun turvallisuus terminologisen tutkimuksen aihealueena .....	17
1.4 Tutkimusaineisto.....	18
1.5 Tutkimusmenetelmät.....	20
1.6 Dispositio .....	20
<b>2 TERMINOLOGIAN PERUSKÄSITTEET .....</b>	<b>22</b>
2.1 <i>Käsitteen</i> käsite terminologian teoriassa.....	23
2.2 Klassisesta luokittelusta elinkaariajatteluun.....	24
2.3 Käsite <i>termi</i> terminologian teoriassa.....	27
2.3.1 Standardinmukainen näkemys termistä .....	27
2.3.2 Ihannetermi sanastotyön oppaiden ja tutkimuskirjallisuuden mukaan.....	29
<b>3 TERMINOLOGIAN TAVOITTEET, TUTKIMUSKOHTEET JA MENETELMÄT.....</b>	<b>32</b>
3.1 Perinteisen terminologian tavoitteet.....	33
3.2 Perinteisen terminologian menetelmät.....	34
3.3 Terminologian tavoitteiden laajeneminen.....	35
3.4 Uudet terminologiset menetelmät.....	36
3.5 Erikoiskielet terminologian näkökulmasta.....	38
<b>4 TERMEN JA KÄSITETIEDON POIMINTAMENETELMÄT.....</b>	<b>43</b>
4.1 Terminpoimintamenetelmät .....	44
4.1.1 Manuaalinen terminpoiminta.....	44
4.1.2 Puoliautomaattinen terminpoiminta.....	45
4.1.2.1 Tilastolliset puoliautomaattiset terminpoimintamenetelmät .....	46
4.1.2.2 Lingvistiset puoliautomaattiset terminpoimintamenetelmät .....	47
4.1.2.3 Puoliautomaattiset yhdistelmämenetelmät terminpoiminnassa .....	48
4.1.2.4 Puoliautomaattisen terminpoiminnan nykytila.....	49
4.2 Käsitetiedon poimintamenetelmät.....	50
<b>5 TERMI TERMINPOIMINNAN NÄKÖKULMASTA .....</b>	<b>54</b>
5.1 Termiesiintymien frekvenssi.....	56
5.2 Termin pituus.....	58
5.3 Termin sanaluokka.....	60
5.4 Saman käsitteen eri nimitykset .....	62
5.5 Konteksti termien ja käsitetiedon poiminnassa .....	65
<b>6 MANUAALINEN JA PUOLIAUTOMAATTINEN TERMINPOIMINTA SUOMALAISISTA JA VENÄLÄISISTÄ MERENKULUN TURVALLISUUTTA KÄSITTELEVISTÄ TEKSTEISTÄ....</b>	<b>68</b>
6.1 Terminpoiminnan aineisto ja menetelmät.....	68
6.2 Terminpoimintatulosten vertailussa käytetyt mittarit.....	72
6.3 Suomalaisen ja venäläisen terminlistan laadinta.....	73
6.3.1 Suomalaisen asiantuntijoiden valitsemien termien ominaisuuksia .....	77
6.3.2 Venäläisten asiantuntijoiden valitsemien termien ominaisuuksia.....	82
6.4 Suomalaisen opiskelijaryhmien terminpoiminta.....	86
6.4.1 Suomalaisen opiskelijaryhmien poimimien termien ominaisuuksia .....	87
6.4.2 Suomalaisen opiskelijaryhmien piilotermin ominaisuuksia.....	90

6.4.3 Suomalaisien opiskelijaryhmien poimimien valetermien ominaisuuksia .....	93
6.5 Venäläisten opiskelijaryhmien terminpoiminta .....	96
6.5.1 Venäläisten opiskelijaryhmien poimimien termien ominaisuuksia .....	96
6.5.2 Venäläisten opiskelijaryhmien piilotermien ominaisuuksia .....	98
6.5.3 Venäläisten opiskelijaryhmien poimimien valetermien ominaisuuksia .....	101
6.6 Erikoisalatiedon ja terminologian tuntemuksen vaikutus terminpoimintaan .....	102
6.7 Puoliautomaattinen terminpoiminta .....	105
6.7.1 Testattavien terminpoimintaohjelmien esittely .....	106
6.7.1.1 Conexor NaviTerm 2.0 .....	106
6.7.1.2 Trados MultiTerm Extract .....	107
6.7.2 Terminpoimintaohjelmien asetukset terminpoimintakokeessa .....	108
6.7.2.1 Lähtetekstin pituuden vaikutus termiehdokkaiden määrään puoliautomaattisessa terminpoiminnassa .....	109
6.7.2.2 Terminpoimintaohjelmien asetusten optimointi .....	110
6.7.3 NaviTerm 2.0 ja Masterin-ohjelmien tuottamat termiehdokaslistat ja asiantuntijalistat vertailu .....	111
6.7.3.1 NaviTerm 2.0 ja Masterin-ohjelmien poimimien termien ominaisuuksia .....	112
6.7.3.2 NaviTerm 2.0 ja Masterin-ohjelmien piilotermien ominaisuuksia .....	115
6.7.3.3 NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien ominaisuuksia .....	117
6.7.4 MultiTerm Extract -ohjelman tuottamien termiehdokaslistojen ja asiantuntijalistojen vertailu .....	120
6.7.4.1 MultiTerm Extract -ohjelman poimimien suomalaisten termien ominaisuuksia .....	121
6.7.4.2 MultiTerm Extract -ohjelman suomalaisten piilotermien ominaisuuksia .....	122
6.7.4.3 MultiTerm Extract -ohjelman poimimien suomalaisten valetermien ominaisuuksia .....	123
6.7.4.4 MultiTerm Extract -ohjelman poimimien venäläisten termien ominaisuuksia .....	124
6.7.4.5 MultiTerm Extract -ohjelman venäläisten piilotermien ominaisuuksia .....	125
6.7.4.6 MultiTerm Extract -ohjelman poimimien venäläisten valetermien ominaisuuksia .....	126
6.8 Terminpoimintamenetelmien vertailu ja arviointi .....	127
6.8.1 Saanti ja tarkkuus terminpoiminnassa .....	128
6.8.2 Termin pituuden vaikutus terminpoimintaan .....	130
6.8.3 Termin frekvenssin vaikutus terminpoimintaan .....	131
6.8.4 Termikaavan vaikutus terminpoimintaan .....	132
6.8.5 Synonymia ja variaatio terminpoiminnassa .....	133
6.8.6 Merenkulun turvallisuuden ydintermit terminpoiminnassa .....	137
6.8.7 Puoliautomaattisen terminpoiminnan etuja ja haittoja .....	139

## **7 KÄSITESUHTEIDEN POIMINTA TERMINOLOGISTEN KOETTIMIEN AVULLA SUOMALAISISTA JA VENÄLÄISISTÄ MERENKULUN TURVALLISUUTTA KÄSITTELEVISTÄ TEKSTEISTÄ.....141**

7.1 Käsitetieto teksteissä .....	141
7.1.1 Käsitesuhteet .....	142
7.1.2 Käsitepiirteet .....	145
7.2 Terminologiset koettimet termien ja käsitetiedon poiminnassa .....	146
7.2.1 Terminologisten koettimien käytön hyödyt .....	146
7.2.2 Terminologisten koettimien käyttöön liittyviä ongelmia .....	147
7.3 Terminologisten koettimien tunnistuksessa käytetty aineisto ja menetelmä .....	148
7.3.1 Tekstikorpusten laajuus .....	149
7.3.2 Tekstien julkaisumuoto .....	151
7.3.3 Tekstilaji .....	151
7.3.4 Viestintätilanne .....	154
7.3.5 Aihealue .....	155
7.3.6 Tyyli .....	156
7.3.7 Tekstien ikä .....	156
7.3.8 Tekstien pituus .....	157
7.3.9 Yhteenveto tutkimusaineistosta .....	157
7.3.10 Terminologisten koettimien tunnistuksessa käytetty menetelmä .....	159
7.4 Suomalaiset terminologiset koettimet .....	161
7.4.1 Käsitteen nimityksen suomalaiset koettimet .....	161
7.4.2 Synonymian suomalaiset koettimet .....	166
7.4.3 Määritelmän suomalaiset koettimet .....	171
7.4.4 Geneerisen käsitesuhteen ylä- ja alakäsitteen suomalaiset koettimet .....	176
7.4.5 Koostumussuhteen ylä- ja alakäsitteen suomalaiset koettimet .....	183
7.4.6 Vieruskäsitteiden suomalaiset koettimet .....	189
7.4.7 Funktiosuhteiden suomalaiset koettimet .....	192
7.4.7.1 Kausaalisuhteiden suomalaiset koettimet .....	192
7.4.7.2 Vaikuttavan tekijän suomalaiset koettimet .....	204
7.4.7.3 Toimintasuhteiden suomalaiset koettimet .....	209
7.5 Venäläiset terminologiset koettimet .....	214



7.5.1 Käsitteen nimityksen venäläiset koettimet .....	215
7.5.2 Synonymian venäläiset koettimet .....	219
7.5.3 Määritelmän venäläiset koettimet .....	221
7.5.4 Geneerisen käsitesuhteen ylä- ja alakäsitteen venäläiset koettimet .....	225
7.5.5 Koostumussuhteen ylä- ja alakäsitteen venäläiset koettimet .....	234
7.5.6 Vieruskäsitteiden venäläiset koettimet .....	241
7.5.7 Funktiosuhteiden venäläiset koettimet .....	244
7.5.7.1 Kausaalisuhteiden venäläiset koettimet .....	244
7.5.7.2 Vaikuttavan tekijän venäläiset koettimet .....	250
7.5.7.3 Toimintasuhteiden venäläiset koettimet .....	254
7.6 Terminologisten koettimien arviointi käsitetiedon poimintamenetelmänä .....	262
7.6.1 Verbit koettimina .....	264
7.6.2 Substantiivit koettimina .....	267
7.6.3 Muut leksikaaliset koettimet .....	269
7.6.4 Ei-leksikaaliset koettimet .....	270
<b>8 JOHTOPÄÄTÖKSET .....</b>	<b>272</b>
<b>LÄHDELUETTELO .....</b>	<b>280</b>
<b>HAKEMISTO .....</b>	<b>293</b>
<b>LIITTEET .....</b>	<b>297</b>

## LYHENTEET

### Terminpoiminta

atfi	suomalaisten asiantuntijoiden ryhmä
atru	venäläisten asiantuntijoiden ryhmä
kofi	suomalaisten kääntämisen opiskelijoiden ryhmä
koru	venäläisten kieliaineiden opiskelijoiden ryhmä
mofi	suomalaisten merenkulun opiskelijoiden ryhmä
moru	venäläisten merenkulun opiskelijoiden ryhmä

### Termikaavat

A	adjektiivi
Abb	lyhenne
Adv	adverbi
C	konjunktio
en	englanninkielinen sana
N	substantiivi
N <sub>c</sub>	yhdyssubstantiivi
N <sub>(c)</sub>	substantiivi tai yhdyssubstantiivi
P	prepositio
S	symboli
V	verbi
X	substantiivilauseke

## KUVIOT JA TAULUKOT

KUVIO 1. Merenkulun turvallisuus on merenkulun osa-alue.....	18
KUVIO 2. Tetraedriksi laajennettu semanttinen kolmio.....	22
KUVIO 3. Käsiteluokkien sijainti aikajanalla elinkaariajattelun mukaan .....	25
KUVIO 4. Erikoiskielten soveltuvuus standardisointiin .....	41
KUVIO 5. Kontekstin kolme tasoa terminologisen tiedon poiminnassa .....	66
KUVIO 6. Merenkulun turvallisuuden käsitekehys ja ydintermit.....	139
TAULUKKO 1. Terminpoiminnassa käytetyn aineiston kuvaus.....	69
TAULUKKO 2. Pilottikokeeseen osallistuneiden merikapteenien termiehdokkaiden poiminta Meriteiden säännöistä.....	74
TAULUKKO 3. Esimerkkejä pilottikokeen koehenkilöiden poimimista erimittaisista termiehdokkaista.....	75
TAULUKKO 4. Suomalaisten asiantuntijoiden poimimien termien pituus sanoina .....	77
TAULUKKO 5. Suomalaisten asiantuntijoiden poimimien termien frekvenssi.....	78
TAULUKKO 6. Suomalaisten asiantuntijoiden poimimien termien yleisimmät termikaavat .....	80
TAULUKKO 7. Venäläisten asiantuntijoiden poimimien termien pituus sanoina .....	83
TAULUKKO 8. Venäläisten asiantuntijoiden poimimien termien frekvenssi.....	84
TAULUKKO 9. Venäläisten asiantuntijoiden poimimien termien yleisimmät termikaavat .....	85
TAULUKKO 10. Suomalaisten asiantuntijoiden ja vähintään neljän koehenkilön poimimien termien pituus.....	88
TAULUKKO 11. Suomalaisten asiantuntijoiden ja vähintään neljän koehenkilön poimimien termien frekvenssi .....	88
TAULUKKO 12. Suomalaisten asiantuntijoiden ja vähintään neljän koehenkilön poimimien termien tavallisimmat termikaavat.....	89
TAULUKKO 13. Suomalaisten koehenkilöryhmien piilotermien pituus .....	90
TAULUKKO 14. Suomalaisten koehenkilöryhmien piilotermien frekvenssi .....	91
TAULUKKO 15. Suomalaisten koehenkilöryhmien piilotermien tavallisimmat termikaavat .....	92
TAULUKKO 16. Suomalaisten koehenkilöryhmien poimimien valetermien pituus ....	93
TAULUKKO 17. Suomalaisten opiskelijaryhmien poimimien valetermien frekvenssi	94
TAULUKKO 18. Suomalaisten opiskelijaryhmien poimimien valetermien tavallisimmat termikaavat .....	94
TAULUKKO 19. Venäläisten asiantuntijoiden ja opiskelijaryhmien poimimien termien pituus .....	97
TAULUKKO 20. Venäläisten asiantuntijoiden ja opiskelijaryhmien poimimien termien frekvenssi.....	97
TAULUKKO 21. Venäläisten asiantuntijoiden ja opiskelijaryhmien poimimien termien tavallisimmat termikaavat.....	98
TAULUKKO 22. Venäläisten koehenkilöryhmien piilotermien pituus.....	99
TAULUKKO 23. Venäläisten koehenkilöryhmien piilotermien frekvenssi.....	99
TAULUKKO 24. Venäläisten koehenkilöryhmien poimimien valetermien pituus.....	101
TAULUKKO 25. Venäläisten koehenkilöryhmien poimimien valetermien frekvenssi .....	102
TAULUKKO 26. Yhteenveto vertailussa käytettyjen termintunnistusohjelmien ominaisuuksista .....	108

TAULUKKO 27. NaviTerm 2.0 -ohjelman poimimien termiehdokkaiden määrän riippuvuus lähtötekstin pituudesta .....	109
TAULUKKO 28. MultiTerm Extract -ohjelman säätöjen vaikutus termintunnistukseen lähtötekstillä pw02 (2 366 sanaa) .....	110
TAULUKKO 29. NaviTerm 2.0 ja Masterin -terminpoimintaohjelmien poimimien termien pituus .....	112
TAULUKKO 30. NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien poimimien termien frekvenssi .....	113
TAULUKKO 31. NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien poimimien termien tavallisimmat termikaavat .....	114
TAULUKKO 32. NaviTerm 2.0 ja Masterin-ohjelmien piilotermien pituus.....	116
TAULUKKO 33. NaviTerm 2.0 ja Masterin-ohjelmien piilotermien frekvenssi .....	116
TAULUKKO 34. NaviTerm 2.0 ja Masterin-ohjelmien piilotermien tavallisimmat termikaavat .....	117
TAULUKKO 35. NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien pituus .....	118
TAULUKKO 36. NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien frekvenssi.....	118
TAULUKKO 37. NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien tavallisimmat termikaavat.....	119
TAULUKKO 38. Terminpoiminnan saanti ja tarkkuus .....	129
TAULUKKO 39. Käsitettä nimeävän ilmauksen terminologiset koettimet suomenkielisessä aineistossa .....	162
TAULUKKO 40. Synonymian terminologiset koettimet suomenkielisessä aineistossa .....	167
TAULUKKO 41. Määritelmän terminologiset koettimet suomenkielisessä aineistossa .....	172
TAULUKKO 42. Geneerisen käsitesuhteen terminologiset koettimet suomenkielisessä aineistossa.....	177
TAULUKKO 43. Koostumussuhteen terminologiset koettimet suomenkielisessä aineistossa.....	184
TAULUKKO 44. Vieruskäsitteiden terminologiset koettimet suomenkielisessä aineistossa.....	190
TAULUKKO 45. Kausaalisuhteen terminologiset koettimet suomenkielisessä aineistossa.....	194
TAULUKKO 46. Vaikuttavan tekijän terminologiset koettimet suomenkielisessä aineistossa.....	205
TAULUKKO 47. Toimintasuhteen terminologiset koettimet suomenkielisessä aineistossa.....	210
TAULUKKO 48. Käsitettä nimeävän ilmauksen terminologiset koettimet venäjänkielisessä aineistossa .....	215
TAULUKKO 49. Synonymian terminologiset koettimet venäjänkielisessä aineistossa .....	219
TAULUKKO 50. Määritelmän terminologiset koettimet venäjänkielisessä aineistossa .....	221
TAULUKKO 51. Geneerisen käsitesuhteen terminologiset koettimet venäjänkielisessä aineistossa.....	226
TAULUKKO 52. Koostumussuhteen terminologiset koettimet venäjänkielisessä aineistossa.....	235

TAULUKKO 53. Vieruskäsitteiden terminologiset koettimet venäjänkielisessä aineistossa.....	241
TAULUKKO 54. Kausaalisuhteen terminologiset koettimet venäjänkielisessä aineistossa.....	245
TAULUKKO 55. Vaikuttavan tekijän terminologiset koettimet venäjänkielisessä aineistossa.....	250
TAULUKKO 56. Toimintasuhteen terminologiset koettimet venäjänkielisessä aineistossa.....	254

## 1 JOHDANTO

Tiedon kerääminen, järjestäminen ja tallentaminen on ihmisille tyypillistä toimintaa. Tyypillistä on myös pyrkimys siirtää tietoa erilaisin menetelmin rajoitetulle tai rajoittamattomalle vastaanottajajoukolle. Nykyisillä tiedonsiirtomenetelmillä tieto saadaan nopeasti monen vastaanottajan ulottuville. Nopea ja tehokas tiedonsiirto on etu, josta emme enää osaisi luopua, sillä sen ansiosta tiedonsaanti reaaliajassa on mahdollista vastaanottajan sijainnista riippumatta. Nopeat ja tehokkaat menetelmät ovat johtaneet maapalloistumiseen, mutta kolikon kääntöpuoli on tiedontulvan ongelma. On arvioitu, että 2000-luvulla maailmassa on jo tuotettu enemmän tietoa kuin koko maailmanhistorian aikana yhteensä, ja tulevaisuudessa tiedon määrän oletetaan kaksinkertaistuvan joka vuosi. Tiedonhallinnasta on tullut keskeinen kysymys niin yritysten, organisaatioiden kuin yksityistenkin ihmisten jokapäiväisessä toiminnassa. Samalla on käynyt selväksi, että tiedon hallinta manuaalisesti ilman kunnollisia työkaluja on ylivoimaista.

Erityisen nopeasti tiedon määrä kasvaa erikoisaloilla, koska yhdessä maassa tuotettu uusi tieto saadaan nopeasti levitettyä maailmanlaajuisesti. Lisääntyvän erikoisalatiedon hallinta on tärkeää kaikille erikoisalan viestintään osallistuville, joita ovat esimerkiksi erikoisalalla toimivien yritysten työntekijät, erikoisalan opettajat ja opiskelijat, tutkijat, tekniset kirjoittajat, kääntäjät, tulkit ja erikoisalan sanastotyötä tekevät terminologit. Nykyisin kaikkien erikoisalojen toimintaympäristö on kansainvälinen, jolloin viestintä on monikielistä ja monikulttuurista. Jotkut viestijöistä, kuten kääntäjät ja tulkit, ovatkin erikoistuneet monikieliseen viestintään.

### 1.1 Terminologinen tutkimus ja sanastotyö tiedon hallinnassa

Tiedonhallinnan ongelmiin on paneuduttu monella tieteenalalla. Erilaisiin tarkoituksiin on kehitetty erilaisia tiedonhallinnan menetelmiä, kuten luokittelujärjestelmiä ja hakumenetelmiä tiedonhakuun sähköisistä lähteistä. Tiedonhallinnan ongelmien ratkaisua haetaan myös terminologisen tutkimuksen keinoin. Terminologisen tutkimuksen yhtenä tavoitteena on nimittäin kehittää työkaluja käytännön sanastotyöhön eli erikoisalojen sanastojen laadintaan. Siihen kuuluu niin käsitelähtöisten ja systemaattisesti järjestettyjen erikoisalanastojen laatiminen kuin yritysten tai kääntäjien ja tulkkien omiin tarpeisiinsa tekemä sanastotyö. Erikoisalan asiantuntijoiden ja terminologioiden yhteisissä sanastoprojekteissa noudatetaan vakiintuneita sanastotyömenetelmiä, ja työn tuloksena syntyneiden sanastojen rakenne on lähellä sanastostandardien rakennetta. Sen sijaan omiin tarpeisiin tehty sanastotyö voidaan toteuttaa monin eri menetelmin, ja valmiiden sanastojen rakenne vaihtelee. Yksinkertaisimmillaan tällainen sanasto on termilista, jossa termeille on esitetty vastineet yhdellä kielellä. Omaan tai organisaation tarkoituksiin laaditun sanaston rakenteen ratkaisee ennen kaikkea sanaston käyttötarkoitus, mutta rakenteeseen vaikuttavat myös käytössä olevat resurssit – sekä taloudelliset että tiedolliset.

Sanastotyön tarkoitus on erikoisalojen viestinnän parantaminen. Siitä hyötyvät esimerkiksi kääntäjät, sillä kääntäjältä kuluu jopa 75 prosenttia käänösajasta termien vastinehakuun (Grinčev 1993: 259). Tärkeää on myös tietää terminmuodostuksen ja käsitteanalyysin periaatteet. Näin voi varmistua siitä, että ehdotettu vastine nimeää juuri sitä käsitettä, josta viestitään. Koska uudet termit päätyvät sanakirjoihin vasta muutaman vuoden kuluttua siitä, kun alan asiantuntijat ovat alkaneet käyttää niitä erikoisalaviestinnässä, uusien tietojen haettavuus erikoisalan teksteistä. Tämä koskee yhtä lailla kääntäjiä, tulkkeja ja teknisiä kirjoittajia kuin erikoisalan opettajia ja opiskelijoitakin.

Historiallisista syistä johtuen sanastotyö oli pitkään normatiivista ja se nähtiin keino-  
nona ohjata erikoisalojen kielenkäyttöä (ks. jakso 3.1). Terminologisessa tutkimukses-  
sakin tutkittavat termit ja käsitetieto haettiin enimmäkseen standardeista ja normatiivi-  
sista sanakirjoista. Viimeksi kuluneiden vuosikymmenien aikana normatiivista sanasto-  
työtä on arvosteltu siitä, ettei siinä oteta huomioon kieleen – myös erikoiskieleen –  
luonnollisesti kuuluvia ilmiöitä, kuten vaihtelua ja muutosta (ks. esim. Temmerman  
2000). Arvostelijoiden mukaan sanastotyössä ja tutkimuksessa on keskitytty ihanneti-  
laan pyrkimiseen ja todellisen kielenkäytön kuvaaminen on jätetty muiden tehtäväksi.  
Termejä ja käsitetietoa on tarkasteltu irrallaan kontekstista, jolloin moni tärkeä ilmiö on  
jäänyt huomaamatta.

Nykyisin normatiivisten sanastojen laatiminen ei enää läheskään aina ole käytännön  
sanastotyön tavoitteena eikä tutkimus kohdistu pelkästään standardien tai normatiivisten  
sanakirjojen termeihin ja määritelmiin. Yhä useammin sekä sanastotyö että tutkimus on  
deskriptiivistä, jolloin kieltä kuvataan sellaisena, kuin sitä käytetään. Terminologiset  
tutkimus- ja sanastotyömenetelmät ovat kehittyneet perinteiseen terminologian teoriaan  
pohjautuvasta ja standardisointia palvelleesta käsiteanalyysistä sosioterminologian me-  
netelmiin, joissa otetaan huomioon erikoiskielen sosiaalinen aspekti sekä käsitteiden ja  
termien elinkaari. Tavallaan on palattu siihen, mistä Vinokur (1939: 5 lähteessä Tata-  
rinov 1994: 30) puhui jo 1930-luvun lopulla: ”Termit eivät ole erityisiä sanoja, vaan  
ainoastaan sanoja erityisessä tehtävässä”. Termejä on siis etsittävä sieltä, missä ne to-  
teuttavat tätä erityistä tehtäväänsä, eli erikoisalaviestinnästä, jota edustavat esimerkiksi  
erikoisalan tekstit.

Teksteistä voidaan löytää erityyppistä terminologisesti kiinnostavaa tietoa. Ensinnä-  
kin teksteistä voidaan hakea käsitteiden nimityksiä, joista terminologisen tiedon poi-  
minnan kannalta tärkeimpiä ovat termit. Terminpoimintaan on olemassa erilaisia mene-  
telmiä: termit voidaan hakea tekstistä manuaalisesti eli käsin poimimalla tai puoliauto-  
maattisesti tietokoneohjelman avulla. Automaattisesta terminpoiminnasta on vielä liian  
aikaista puhua, sillä puoliautomaattiset menetelmät eivät kehitystyöstä huolimatta ole  
pystyneet täysin poistamaan manuaalisen työn tarvetta. Puoliautomaattisessa termin-  
poiminnassa onkin itse asiassa kyse yhdistelmämenetelmästä, jossa termiehdokkaat  
poimitaan ensin automaattisesti tietokoneohjelman avulla ja poiminnan tuloksena saatu  
termiehdokaslista tarkistetaan ja muokataan manuaalisesti. Toiseksi teksteistä voidaan  
hakea käsitetietoa eli tietoa käsitesuhteista ja käsitepiirteistä. Käsitetiedon avulla pyri-  
tään selvittämään käsitteiden väliset suhteet. Termien ja käsitetiedon lisäksi teksteissä  
saattaa olla muutakin terminologisesti kiinnostavaa tietoa, kuten ensyklopedista tietoa  
käsitteen historiasta tai toiminnasta, joka jää perinteisillä menetelmillä poimimatta, kos-  
ka se ei sovi ennalta annettuun malliin. Tieto käsitteen historiasta tai toiminnasta ei ehkä  
ole välttämätöntä määritelmän muotoilussa tai vastinehaussa, mutta joskus lisätieto kä-  
sitteestä saattaa olla tarpeen.

Tarvittava tieto on nykyisin poimittava kasvavasta erikoisalan tietomassasta mahdol-  
lisimman nopeasti. Valtavan tietomassan järjestäminen on kuitenkin haastavaa, ja tämän  
vuoksi terminologiselle tutkimuksellekin on asetettu uusia tavoitteita, sillä perinteiset  
tutkimus- ja sanastotyömenetelmät eivät enää riitä. Perinteiset käsityövaltaiset mene-  
telmät ovat liian hitaita ja epätarkkojakin, kun aineiston koko kasvaa satojen tai tuhansi-  
en sivujen laajuiseksi. Terminologian tutkijat ovatkin pyrkineet kehittämään automaatti-  
sia tai ainakin puoliautomaattisia menetelmiä termien ja käsitetiedon hakuun ja poimin-  
taan suuresta tekstimassasta. Terminologian tutkijat ja sanastotyön tekijät ovat kuitenkin  
uusien menetelmien kehittämisessä ja soveltamisessa jäljessä, sillä lähialoilla, kuten  
tietokonelingvistiikassa, informaatiotieteissä ja korpuslingvistiikassa automaattisten ja  
puoliautomaattisten menetelmien kehittämisessä ollaan paljon pitemmällä. Menetelmien

lainaaminen lähialoilta on houkuttavaa, mutta näiden alojen menetelmät eivät useinkaan sellaisenaan sovi terminologisen tutkimuksen ja sanastotyön tarpeisiin, koska tavoitteet ovat erilaiset. Terminologian tutkijatkaan eivät välttämättä ole tiedostaneet tavoitteiden eroa, vaan ovat ottaneet käyttöön lähialojen menetelmiä, joilla termiehdokkaita poimitaan teksteistä erillään käsitetiedosta. Näin menetetään runsaasti arvokasta tietoa muun muassa käsitesuhteista ja käsitepiirteistä. Tämä tieto on hankittava muuten, vaikka suuri osa siitä olisi teksteissä valmiina. Terminpoiminnan sijasta kannattaisikin keskittyä terminologisen tiedon poimintaan, jolloin termit ja muu käsitetieto poimitaan yhdessä.

Perinteisessä sanastotyössä termiehdokkaiden poimintaa seuraavat työvaiheet ovat käsitteiden valinta, käsitesuhteiden selvittäminen ja määritelmien muodostaminen sanastoon valituille käsitteille. Näissä työvaiheissa yleisesti käytetty menetelmä on perinteinen käsiteanalyysi, jonka tavoitteena on muodostaa lähikäsitteistä hierarkkinen geenierinen tai koostumussuhteinen käsitejärjestelmä, jossa käsitteet on määritelty olennaisten ja riittävien käsitepiirteiden avulla. Käsitejärjestelmän rakentamiseen ja määritelmien muodostamiseen tarvittava tieto on yleensä hankittu erikoisalan asiantuntijoilta tai teksteistä manuaalisesti. Tekstejä ei ole pidetty erityisen hyvinä käsitetiedon lähteenä, sillä niissä ei käsitteille aina ole suoraan annettu olennaisia ja riittäviä käsitepiirteitä, joita tarvitaan määritelmän muodostamiseen. Suhtautuminen teksteihin käsitetiedon lähteenä on kuitenkin muuttumassa, sillä nyt ymmärretään, miten hyvin teksteistä löytyy myös käsitetietoa. Terminologisen tutkimuksen ja sanastotyön tekijöiden mielenkiinto suuntautuukin nykyisin yhä enemmän todelliseen kielenkäyttöön. (Ks. esim. Pearson 1998.)

Kiinnostuksen kohteen muutos tuo eteen uusia haasteita, sillä todellisen kielenkäytön kuvaamiseen tarvitaan perinteisen käsiteanalyysin lisäksi muitakin menetelmiä. Moni tutkija on todennut, että teksteissä on kielellisiä ja ei-kielellisiä signaaleja, jotka ilmaisevat muun muassa synonymiaa, käsitesuhteita ja käsitepiirteitä. Signaalit voivat olla sanoja, ilmauksia, painoteknisiä keinoja, välimerkkejä tai kuvia. On tärkeä tuntea nämä signaalit ja tulkita ne oikein. Kielentutkijat ovat aiemminkin käsitelleet erilaista kielellistä tietoa ilmaisevia signaaleja, mutta tietotekniikan kehitys on vauhdittanut menetelmän tutkimusta.

Tekstiaineistojen analyysimenetelmien kehittymisestä huolimatta termien ja käsitetiedon poimintaan teksteistä ei ole tyydyttäviä, kielestä riippumattomia menetelmiä, vaan tämä aikaa vievä tutkimuksen tai sanastotyön vaihe tehdään yhä usein manuaalisesti ja intuitioon perustuen. Varsinkin pienten kielten ongelma on, että sekä automaattisten termintunnistusmenetelmien että käsitetiedon poimintamenetelmien kehitystyötä ja tutkimusta on toistaiseksi tehty lähinnä englanninkielisillä aineistoilla. Englannin kielen valinta on ymmärrettävää, sillä englanti on kansainvälinen tieteen kieli ja suuri osa erikoisalojen teksteistä on kirjoitettu englanniksi. Suomen- tai venäjänkielisellä aineistolla ei uusien terminologisten menetelmien soveltuvuutta ole juurikaan tutkittu, vaikka näiden kielten termien tunnistukseen soveltuvia termintunnistusohjelmia on kaupallisesti tarjolla.

## 1.2 Työn tarkoitus

Edellä olen muotoillut terminologian tavoitteeksi erikoisalatiedon hallinnan ja sitä kautta erikoisalaviestinnän parantamisen. Maapalloistuneessa monikielisessä erikoisalan yhteisössä erikoisalatiedon hallinta on ainakin toistaiseksi monikielistä ja monikielisestä viestinnästä vastaavien on osattava hakea termeille ja muille erikoisalan käsitteiden nimityksille oikeat vastineet. Tämän tutkimuksen tavoitteena on tuottaa tietoa sellaisten työkalujen kehittämiseksi, joiden avulla aidoista erikoisalateksteistä koostuvasta aineis-



tosta voidaan paikallistaa käsitteet sekä poimia termit ja mahdollinen käsitetieto eli tieto käsitesuhteista ja käsitepiirteistä. Tällä tavoin työ välillisesti edesauttaa myös monikielistä viestintää, sillä käsitetiedon paikallistaminen on tarpeen, jotta erikielisiä käsitteitä voidaan verrata ja näin selvittää, missä määrin ja missä suhteissa ne vastaavat toisiaan. Tästä syystä tutkimuksessa on kaksi kieltä, suomi ja venäjä, vaikka varsinaiseen vastavuoksien selvittämiseen ei työssäni ole mahdollisuutta.

Ensimmäisenä osatavoitteena on tutkia, millaisia termiehdokkaita nykyisin käytössä olevat terminpoimintamenetelmät tuottavat. Tarkastelussa ovat manuaalinen terminpoimintamenetelmä, jossa termit poimitaan käsin, ja puoliautomaattinen menetelmä, jossa terminpoimintaan käytetään tietokoneohjelmaa. Erilaisilla menetelmillä tuotettuja termiehdokaslistoja vertaamalla pyrin selvittämään, 1) vaikuttaako erikoisalan tuntemus manuaalisen terminpoiminnan tulokseen, 2) millaisia ovat koetekstin keskeiset termit, jotka eivät tuota vaikeuksia terminpoiminnassa, 3) millaiset termit tuottavat vaikeuksia manuaalisessa ja puoliautomaattisessa terminpoiminnassa ja 4) kuinka käyttökelpoisia ovat kaupallisten terminpoimintaohjelmien tuottamat termiehdokaslistat. Tavoitteeni on tuottaa tietoa terminpoimintaan tarkoitettujen työkalujen kehittämiseksi, ei terminpoiminnan käytännön ongelmien ratkaisemiseksi.

Toisena osatavoitteena on selvittää, millä tavoin tekstistä voidaan poimia käsitetietoa sähköisten apuvälineiden avulla ja miten käsitetiedon poiminnan voi yhdistää terminpoimintaan. Käsitetiedon poimintaa yhdistyneenä termien poimintaan ei ole juuri tutkittu, sillä termejä on usein tarkasteltu irrallaan niistä käsitteistä, joita ne nimeävät. Tässä tutkimuksessa pyrin kuitenkin säilyttämään yhteyden termien ja käsitetiedon välillä. Tähän tavoitteeseen pyrin tutkimalla, soveltuvatko terminologiset koettimet (knowledge probes) termien ja käsitetiedon poimintaan suomen- ja venäjänkielisistä erikoisalan teksteistä. Terminologisilla koettimilla tarkoitan kielellisiä ilmauksia, painoteknisiä keinoja tai välimerkkejä, jotka ilmaisevat termien tai käsitteiden välisiä suhteita tai käsitepiirteitä. Terminologiset koettimet ovat keskeisesti esillä tutkimuksessani. Siksi halusin niiden näkyvän työni otsikossa *Merenkulun turvallisuuden koetinkiviä*, jossa käytän sanaa *koetinkivi* metaforisesti.

Lopullisena tavoitteena on luoda pohjaa erityisesti kääntäjän tarpeisiin soveltuvan termien ja käsitetiedon poimintamenetelmän kehittämiseksi. Tämä menetelmä ei vaadi laajaa korpusta tai aineiston valmistelevaa käsittelyä, koska esimerkiksi itsenäisellä kääntäjällä ei todennäköisesti ole käytössään laajaa erikoisalakorpusta eikä kalliita tietoteknisiä välineitä korpusten käsittelyyn. Tästä näkökulmasta tutkimukseni ensisijaisena kohderyhmänä ovat terminologisen tiedon hallintaan tarkoitettujen työkalujen kehittäjät.

Erikoisalan käsitteitä nimeävät termit ja käsitteiden määritelmät kiinnostavat laajaa kieliammattilaisten joukkoa, joille ajantasainen erikoisalatieto on elintärkeää. Tutkimuksesta on hyötyä terminologisten sanastojen ja termipankkien laatijoille, teknisille kirjoittajille, kääntäjille ja tulkeille, joiden tulee 1) paikallistaa tekstistä termit, 2) tunnistaa käsitteiden väliset suhteet ja 3) poimia käsitepiirteitä teksteistä käsitevastaavuuksien selvittämiseksi. Tutkimuksessa pyritään kehittämään menetelmä, joka sopii sekä lähtötekstin analysointiin että samaa aihetta käsittelevien erikielisten tekstien analysointiin vastineiden haussa. Etukäteen on selvää, että ehjän käsitejärjestelmän rakentaminen teksteistä poimitun tiedon avulla on käytännössä mahdotonta, koska aineistoa ei voi koostaa niin, että siinä olisi kaikki rajatun erikoisalan käsitteet ja niiden kaikki olennaispiirteet. Teksteistä saadun tiedon avulla voidaan kuitenkin rakentaa alustava käsitejärjestelmä, jota voi täydentää alan asiantuntijoita haastatteleamalla. Tämä sanastotyön vaihe jää kuitenkin tämän tutkimuksen ulkopuolelle. Kääntäjät ja terminologit hyötyvät tutkimuksesta myös välillisesti, kun tutkimustuloksia käytetään entistä käyttökelpoisempien tietotekniikka-apuvälineiden kehittämiseen.

Tutkimuksesta on hyötyä myös erikoisalan asiantuntijoille, jotka osallistuvat oman alansa viestintään esimerkiksi kirjoittamalla tai luennoimalla. Heidän on kehitettävä, uudistettava ja muotoiltava termejä, laadittava määritelmiä sekä luokiteltava ja nimettävä käsitteitä. Tässä tutkimuksessa kuvattujen menetelmien soveltaminen auttaa erikoisalan asiantuntijoita näissä tehtävissä. Lisäksi tutkimuksesta on hyötyä terminologian opetukselle, erikoisalojen opetukselle ja oppimateriaalin laatijoille.

Esittelen tutkimuksessani erilaisilla terminpoimintamenetelmillä tuotettujen termiehdokkaiden ominaisuuksia ja yhden tavan poimia käsitetietoa teksteistä. Termistöjä voi koota ja käsitetietoa hankkia muillakin menetelmillä, esimerkiksi asiantuntijahaastatteluilla, mutta olen rajannut näiden menetelmien esittelyn tutkimukseni ulkopuolelle. Tarkasteltavien menetelmien valintaan vaikutti erityisesti se, että tutkimukseni tuottamaa tietoa voidaan käyttää hyväksi kehitystyössä, jonka tavoitteena on terminpoiminnan ja käsitetiedon poiminnan automatisointi.

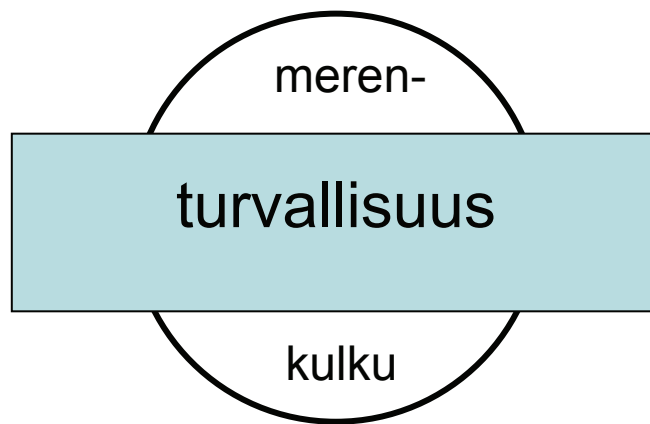
### **1.3 Merenkulun turvallisuus terminologisen tutkimuksen aihealueena**

Terminologisen tutkimuksen kohteena ovat erikoisalojen käsitteet ja termit. Tästä syystä terminologisen tutkimuksen aineisto on koottava tietyltä tai tietyiltä erikoisaloilta. Tässä tutkimuksessa aineisto on koottu merenkulun erikoisalalta. Koska merenkulku on hyvin laaja ala, aihealue on rajattu merenkulun turvallisuuteen. Erikoisalan valintaan ja rajaukseen vaikuttivat alan tärkeys ja kansainvälisyys sekä myös alalta tehdyn terminologisen tutkimuksen vähäisyys.

Merenkulku on tärkeä elinkeino niin Suomessa kuin Venäjälläkin. Suomessa ulkomaankaupan kuljetuksista noin 80 prosenttia tapahtuu meritse. Meriturvallisuus on viime vuosina ollut usein esillä tiedotusvälineissä ja korkean tason tapaamisissa. Kansainvälisyys on aina ollut merenkululle tyypillistä, mutta sekamiehistöjen yleistyessä kansainvälisyys korostuu entisestään. Itse asiassa on yllättävää, miten vähän merenkulun kieltä on tutkittu terminologisin menetelmin, vaikka sekä meri- että lentoliikenteessä on tapahtunut onnettomuuksia, joissa kieli on ollut yksi onnettomuuteen johtanut tekijä. Muualla kuin Suomessa merenkulun kieli on toki ollut tutkimuksen kohteena aiemminkin, mutta tutkimus on usein kohdistunut englanninkielisiin sanakirjoista poimituihin termeihin eikä siihen, miten merenkulun käsitteitä kuvataan tai nimetään autenttisissa teksteissä (ks. esim. Bondarenko 1992; Opitz 1979; Pritchard 2003).

Erikoisalan rajaaminen on aina vaikeaa, sillä erikoisalat limittyvät. Erikoisalojen viestinnässä käytettävien erikoiskielten rajaaminen ei siis voi olla ainakaan helpompaa. Kuten Ahmad ja Salway (1997: 95–96) sanovat, esimerkiksi tieteen kieli tai lain ja hallinnon kieli ovat tietyn enemmän tai vähemmän selvärajaisen erikoisalan kieliä kun taas esimerkiksi opetuskieli ja myyntikieli ovat monelle erikoisalalle ulottuvia yhteiskunnallisen tai inhimillisen toiminnan kieliä. Tämän tutkimuksen aihealueen rajaus merenkulun turvallisuuteen antoi periaatteessa kaksi mahdollista näkökulmaa. Aihealuetta voi tarkastella turvallisuusnäkökulmasta yhtenä turvallisuuden osa-alueena, jolloin merenkulun turvallisuuden kieli on turvallisuuskielen osa. Turvallisuuskieli on monelle alalle ulottuva kieli, jolla viestitään turvallisuusasioista eri aloilla. Turvallisuusnäkökulmasta tarkasteltuna merenkulku sijoittuu laajempaan liikenteen kontekstiin, johon meriliikenteen lisäksi kuuluvat ilma- ja maaliikenne. Merenkulun turvallisuutta ja turvallisuusjärjestelmiä onkin usein verrattu ilmailun mutta myös teollisuuslaitosten turvallisuuteen. Toinen mahdollisuus on tarkastella merenkulun turvallisuutta yhtenä merenkulun osa-alueena, jolloin merenkulun turvallisuuden kieli on merenkulun kielen osa. Merenkulun näkökulmasta merenkulun turvallisuutta on käytännössä vaikea erottaa erilliseksi osa-alueeksi, koska turvallisuus on mukana kaikessa merenkulkuun liittyvässä toiminnassa,

kuten laivanrakennuksessa, navigoinnissa, lastinkäsittelyssä tai aluksen huoltotöissä. Merenkulun turvallisuuden kieli voidaan kuitenkin karkeasti määritellä merenkulun kielen osaksi, jota tarvitaan merenkulun turvallisuusasioista viestimiseen.



KUVIO 1. Merenkulun turvallisuus on merenkulun osa-alue

Tässä tutkimuksessa lähtökohtana on nimenomaan merenkulun turvallisuus, ei turvallisuus yleensä. Aineisto ei siten koostu esimerkiksi työturvallisuusmääräyksistä tai varoituskilvistä, vaan teksteistä, joissa käsitellään aluksen turvallista kulkua. Aihealuetta on rajattu muillakin tavoin. Eri alustyypeistä tässä työssä käsitellään vain kauppa-aluksia ja niistäkin vain kuivarahtialuksia ja tankkereita. Tutkimusalueen ulkopuolelle jäävät siten esimerkiksi matkustaja-alukset ja sota-alukset. Olen kaventanut alueen rajausta myös siten, että käsitelen vain rahtialuksen käyttöön liittyviä turvallisuusasioita, en alusten rakenteellisia ominaisuuksia, vaikka ne kuuluvat *merenkulun turvallisuus* -käsitteen alaan, koska rakenteellisiin turvallisuusominaisuuksiin voi vaikuttaa vain rakennusvaiheessa tai telakoinnin yhteydessä.

Tutkimuksessani keskityn komentosiltatyöskentelyyn mitenkään väheksymättä muiden osastojen henkilökunnan osuutta. Merenkulun turvallisuus riippuu aluksen koko henkilökunnasta, mutta suurimman vastuun aluksen turvallisesta kulusta kantaa kuitenkin kansipäällystö.

#### 1.4 Tutkimusaineisto

Terminologisissa tutkimuksissa aineistona voidaan käyttää 1) valmiita sanakirja- ja sanastoaineistoja (secondary texts), 2) tekstejä tai 3) yhdistelmäaineistoja, joissa on sanakirjoja, sanastoja ja tekstejä. Aiemmissa tutkimuksissa on yleensä keskitytty sanakirjoissa esiintyviin termeihin ja määritelmiin (ks. esim. Pilke 2000). Tämä on ymmärrettävää, sillä sanakirjat ovat helppokäyttöisiä, koska tieto on niissä valmiiksi järjestetty ja käsitetieto muokattu määritelmämuotoon. Tutkijan ei tarvitse pohtia esimerkiksi kysymystä, mikä on termi. Osa tutkijoista on käyttänyt aineistona tekstejä (esim. Kageura ym. 2000; Karihalme 1996) tai sanakirjojen ja tekstien yhdistelmää (esim. Guljakova 1984). Tutkimusaineiston koostamisperusteisiin vaikuttaa ennen muuta tutkimuksen tavoite. Jos valmiiden sanakirjojen käyttö on tavoitteen kannalta tarkoituksenmukaista, tutkija säästyy tekstikorpuksen kokoamisen vaivalta. Sanakirjoissa ei kuitenkaan usein ole uusimpia erikoisantermejä, eivätkä ne kuvaa erikoiskielen todellista käyttöä, sillä

sanakirjoissa termit ja määritelmät on irrotettu niiden luonnollisesta käyttöympäristöstä erikoisalan viestinnässä.

Nykyisin sanastotyö on yhä useammin pikemmin deskriptiivistä kuin standardisointiin pyrkivää. Deskriptiivisen erikoisalan sanakirjan tai -sanaston tarkoitus on nimensä mukaisesti kuvata erikoisalan todellista kielenkäyttöä. Tästä syystä erikoisalan viestinnässä käytetty kieli, joko kirjoitettu tai puhuttu, on sanastotyön tekijän ensisijainen aineistolähde. Deskriptiivisessä sanastotyössä aineiston koostamisen ja käsittelyn osuus on ajallisesti ja työmäärällä mitattuna merkittävä. Tutkijoiden tehtävä onkin kehittää menetelmiä, jotka helpottavat ja tehostavat aineiston koostamista ja käsittelyä.

Näistä käytännön sanastotyön tarpeista on noussut kiinnostus käyttää tutkimusaineistona tekstejä tai teksti- ja sanakirja-aineistojen yhdistelmiä pelkästään sanakirjoista ja sanastoista koostuvien aineistojen sijasta. Tietotekniikan kehityksen ansiosta tekstiaineistojen käyttö on entistä helpompaa, ja osa tutkijoista onkin käyttänyt tutkimusmateriaalina tekstejä (ks. esim. Karihalme 1996; Pearson 1998; Temmerman 2000). Samasta syystä tekstien käyttö aineistona yleistyy sanastotyötä tekevien terminologien ja kääntäjien keskuudessa. Tekstien etu on, että niissä on kontekstitietoa, joka on tarpeen esimerkiksi kääntäjille.

Koska tutkimukseni tavoitteena on kehittää termien ja käsitetiedon poimintamenetelmiä esimerkiksi deskriptiivisen sanastotyön tarpeisiin, aineistoni koostuu teksteistä. Tekstillä tarkoitan tässä tutkimuksessa jotakin viestintätehtävää varten kirjoitettua pitempää kokonaisuutta, esimerkiksi aikakauslehtiartikkelia tai artikkelin muotoon kirjoitettua esitelmää. Erotukseksi sanastoaineistoista, joille on ominaista jonkinasteinen kiivettyneisyys ja keinotekoisuus, voi aineistoni sanoa koostuvan aidoista käyttöteksteistä. Olen valinnut aineistooni aitoja käyttötekstejä erityisesti siksi, että niissä puhutaan erikoisalalle vastikään tulleista käsitteistä. Teksteissä on todennäköisesti myös uusimmat termit, joita ei sanastoissa vielä ole.

Olen valinnut aineistooni sellaisia tekstejä, jotka voisivat olla myös tutkimustulosteni sovelluskohteina. Tutkimustulosten todennäköisten käyttäjien eli terminologien, kääntäjien, tulkkien, teknisten kirjoittajien ja erikoisalan opettajien työssä erikoisalan tekstit ovat todennäköisesti ennen kaikkea säädöksiä, sopimuksia ja viranomaisasiakirjoja, mutta myös tutkimusraportteja, artikkeleita ja oppikirjoja. Aineiston valinnan yhteydessä nousi esiin kysymys säädöstekstien soveltuvuudesta aineistoon. Koska kansainväliset merenkulun turvallisuutta käsittelevät säädökset ja sopimukset on alun perin kirjoitettu englanniksi ja myöhemmin käännetty muille kielille, esimerkiksi suomeksi ja venäjäksi, tuntui houkuttevalta valita aineistoon säädös- ja sopimuskäännöksiä. Houkutus oli varsin suuri, sillä alkuperäistä englanninkielistä säädöstekstiä olisi voinut käyttää englanninkielisten terminologisten koettimien automaattisessa haussa. Vastaavat suomen- ja venäjänkieliset ilmaukset olisi näin ollut suhteellisen helppo poimia säädöskäännöksistä. Päädyin kuitenkin toisenlaiseen ratkaisuun, sillä tarkoituksena on poimia aitoja suomen- ja venäjänkielisiä koettimia eikä käännösvastineita. Vaikka kääntäjät todennäköisesti ovat käyttäneet enimmäkseen samoja ilmauksia kuin suomalaiset tai venäläiset kirjoittajat käyttävät omalla äidinkielellään kirjoittaessaan, käännösinterferenssin mahdollista vaikutusta ei voi sivuuttaa. Käännöstekstit sopivat käännöstutkimukseen, mutta kielen ilmiöiden tutkimuksessa alkukieliset tekstit ovat parhaita. Näistä syistä olen jättänyt säädöstekstit kokonaan aineiston ulkopuolelle. Alkukielisten tekstien tutkimus on käännöstieteessäkin noussut käännösten tutkimuksen rinnalle, sillä alkukielisistä teksteistä kielen yksiköitä ja piirteitä voidaan tutkia niiden kotiympäristössä (home environment; Baker 1995: 232). Mikään ei kuitenkaan estä käyttämästä tässä tutkimuksessa kuvattua termien ja käsitetiedon poimintamenetelmää esimerkiksi säädöskäsitteistöön.

Tutkimusaineiston valinnan lähtökohtana on ollut yksi ydinkäsite, *merenkulun turvallisuus*. Aineistohaussa olen käyttänyt lähtökohtana kansainvälistä kirjastojen UDK-luokitusta, jossa merenkulku kuuluu luokkaan 656.6. Kirjastoluokittelun ulkopuolinen aineisto on valittu lähinnä tekstin otsikon perusteella. Tutkimusaineisto koostuu 1990- ja 2000-luvulla julkaistuista suomen- tai venäjänkielisistä oppikirjoista, tutkimusraporteista, aikakauslehtiartikkeleista ja kirjoitetuista esitelmistä (ks. Lähdeluettelo: Aineslähteet).

Aineiston valintaan vaikuttivat myös valitut tutkimusmenetelmät. Koska tutkimuksessa on käytetty manuaalista menetelmää sekä terminpoiminnan että terminologisten koettimien tarkastelussa, aineiston määrä oli rajattava sellaiseksi, että sen manuaalinen käsittely kohtuullisessa ajassa oli mahdollista. Suomenkielisen aineiston laajuus on 74 215 ja venäjänkielisen 41 025 sanaa, eli laajuus on yhteensä 115 240 sanaa. Aineisto on korpukseksi varsin pieni, mutta tämän tutkimuksen tarpeisiin riittävä, sillä tarkoitus on tutkia menetelmien toimivuutta eikä viedä läpi laajaa sanastotyöprojektia. Korpus on koottu kirjastojen ja yksityisten henkilöiden arkistoista ainoastaan tätä työtä varten, sillä tutkimuksen tarkoituksiin sopivaa valmista korpusta ei ollut käytettävissä. Koottu suomen- ja venäjänkielinen merenkulun turvallisuuden tekstiaineisto onkin toistaiseksi luultavasti ainut tällainen aineisto. Tutkimusaineiston tarkempi kuvaus on jaksossa 7.3.

## 1.5 Tutkimusmenetelmät

Tutkimukseni on laadullinen menetelmien soveltuvuuden tutkimus. Sen empiirisen osuuden alussa sovellan kokeellisesti opiskelijoista muodostettujen koehenkilöryhmien tekemää manuaalista ja tietokoneohjelman avulla tapahtuvaa puoliautomaattista menetelmää termien poimintaan. Opiskelijaryhmien ja tietokoneohjelmien tuottamia termiehdokaslistoja vertaan terminologien ja merenkulun ammattilaisten yhteistyössä koostamaan termilistään. Tutkimuksen tässä osassa on tavoitteena selvittää, onko erikoisalan koulutuksella vaikutusta terminpoimintaan ja kuinka hyvin kokeeseen valittujen kaupallisten terminpoimintaohjelmien tuottamat termiehdokaslistat vastaavat terminologien ja merenkulun asiantuntijoiden tuottamia termilistoja. Terminologisen tutkimuksen uusimman suuntauksen korostama sosiaalinen ulottuvuus on pyritty ottamaan huomioon käyttämällä aineistona aitoja käyttötekstejä sanastojen sijaan.

Käsitteiden paikallistamiseen ja käsitetiedon poimintaan teksteistä kokeilen terminologisia koettimia. Tutkimuksen tämän osuuden tavoitteena on selvittää, soveltuvatko terminologiset koettimet käsitesuhteiden poimintaan suomen- ja venäjänkielisistä erikoisalan teksteistä. Niiden tutkimiseen käytän manuaalisen ja puoliautomaattisen menetelmän yhdistelmää. Kokoan tutkimusaineistosta manuaalisesti kaikki sellaiset ilmaukset tai graafiset merkit, jotka toimivat koettimina. Tämän jälkeen haen aineistosta korpustyökalun avulla kaikki näiden ilmausten ja merkkien esiintymät. Korpustyökalun konkordanssitoiminnon tuottamasta koettimien ja ympäröivien tekstikontekstien listasta poimin ne kontekstit, joissa koetin todella ilmaisee terminologista tietoa. Tällä menetelmällä saadaan koottua alustava luettelo suomen- ja venäjänkielistä koettimista, joiden avulla teksteistä voidaan poimia ainakin osa käsitteiden välisiin suhteisiin liittyvästä tiedosta. Näin tutkimus pohjustaa käytännön sanastotyön automatisointia.

## 1.6 Dispositio

Luvut 2, 3 ja 4 muodostavat työn teoreettisen osan. Työn teoreettisen viitekehyksen esittelen luvuissa 2 ja 3. Luvussa 2 käsittelen tutkimukseni kannalta tärkeimmät terminologian peruskäsitteet eli käsitteen ja termin. Luvussa 3 selostan terminologian ta-

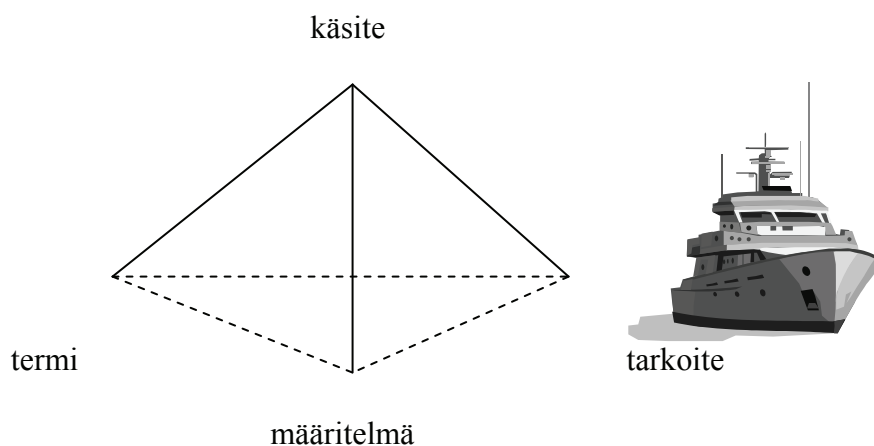
voitteita, tutkimuskohdetta ja menetelmiä. Kuvaan luvussa yleisesti terminologian kenttää alkaen normatiivisesta terminologiasta ja edeten kohti nykyistä deskriptiivisempää lähestymistapaa ja kognitiivisia suuntauksia. Luvussa 4 kuvaan työssä tarkasteltavat termien ja käsitetiedon poimintamenetelmät, joita ovat manuaalinen ja sähköisten apuvälineiden avulla tehtävä puoliautomaattinen poiminta.

Lukuihin 5, 6 ja 7 on koottu työn empiirinen osa. Luvussa 5 tarkastelen terminpoimintaan vaikuttavia määrällisiä ja kielellisiä termin ominaisuuksia. Luvussa 6 selostan terminpoimintakoetta, jossa sovelsin valittuja manuaalisia ja puoliautomaattisia terminpoimintamenetelmiä suomalaisten ja venäläisten merenkulun turvallisuustermien poimintaan teksteistä. Luvussa 7 kuvaan teoreettisesti käsitesuhteet, käsitepiirteet ja terminologisten koettimien käytön käsitetiedon poiminnassa. Tässä luvussa kerron myös tutkimusaineiston koostamisperiaatteista yleensä ja erityisesti tämän tutkimuksen aineiston koostamisesta. Tämän jälkeen selostan käsitetiedon poimintakoetta, jossa hain osittain manuaalisesti ja osittain korpusohjelman avulla käsitesuhteiden poimintaan sopivia suomalaisia ja venäläisiä terminologisia koettimia merenkulun turvallisuutta käsittelevistä suomen- ja venäjänkielisistä teksteistä. Samalla luon teoreettista pohjaa käsitesuhteiden tunnistamisen osittaiselle automatisoinnille. Työn viimeisessä osassa, luvussa 8, esitän tutkimukseni johtopäätökset.



## 2 TERMINOLOGIAN PERUSKÄSITTEET

Terminologian oppikirjoissa terminologian teoriaan tutustuminen aloitetaan yleensä semantiikan alalta lainatusta semanttisesta kolmiosta, niin sanotusta Ogdenin ja Richardsin kolmiosta, jonka kärkeen on sijoitettu semantiikan käsitteitä vastaavat terminologian peruskäsitteet *käsite*, *termi* ja *tarkoite*. Kolmiomalli on terminologisessa kirjallisuudessa laajennettu tetraedriksi, jonka neljänteen kärkeen on sijoitettu *määritelmä*.



KUVIO 2. Tetraedriksi laajennettu semanttinen kolmio

Tetraedrin huipulle sijoitetaan yleensä *käsitteen* käsite, koska sitä pidetään terminologian lähtökohtana. Samasta syystä peruskäsitteiden kuvaus aloitetaan käsitteestä. Määritelmän mukaan käsite on tietyistä käsitepiirteistä muodostuva tiedon yksikkö (Terminologian sanasto, s.v. *käsite*). Jaksossa 2.1 tarkastelen *käsitettä* teoreettisesti siinä laajuudessa, kuin se tämän tutkimuksen kannalta on tarpeen.

Jotta erikoisalan käsitteestä voidaan viestiä, se nimetään kielellisellä ilmauksella, joka saattaa aluksi olla pitkä, jopa lähellä määritelmää. Vähitellen se lyhenee ja sen käyttö vakiintuu, jolloin sitä voidaan kutsua termiksi. Jaksossa 2.3 selostan *termin* käsitettä.

Tässä tutkimuksessa en paneudu tarkemmin tetraedrin kolmanteen kärkeen sijoittuvan *määritelmän* käsitteeseen, vaikka käsittelenkin jaksoissa 7.1.1 ja 7.1.2 määritelmän muotoilussa tarpeellista tietoa, kuten tietoa käsitesuhteista ja käsitepiirteistä. Käytän tästä tiedosta nimitystä *käsitetieto*. Mainitut jaksot antavat taustatietoa tutkimukseni empiiriselle osalle, jossa selostan termien ja käsitetiedon poimintaa teksteistä.

Terminologian keskeisten käsitteiden suhteita kuvaavan tetraedrin neljänteen kärkeen on sijoitettu termi *tarkoite*, jolla viitataan reaali maailman esineeseen, ilmiöön tms. En kuitenkaan pohdi tarkemmin tarkoitteen ja käsitteen suhdetta, koska tarkoitteen tutkimus ei varsinaisesti kuulu terminologian alaan, vaikka terminologiassa on olemassa myös sellainen menetelmä, jossa termiä ja tarkoitetta tutkitaan käsi kädessä. Menetelmän englanninkielinen nimitys on *word-and-object research* ja saksankielinen *Wörter und Sachen Schule*. Menetelmän käyttöä puolustaa ennen kaikkea se, että siinä yhdistyvät synkroninen termintutkimus ja diakroninen historiallis-etymologinen tutkimus. Menetelmää on menestyksekkäästi sovellettu muun muassa suomenruotsalaiseen hylkeenpyynnin termistöön (Nyström 2000). Menetelmän puutteena voidaan pitää keskittymistä termin ja tarkoitteen suhteeseen eikä erikoisalatietoon käsite lähtökohtana (vrt. Pilke 2002: 10).

Aloitan terminologian peruskäsitteiden tarkastelun *käsitteen* käsitteestä, koska sitä pidetään terminologian lähtökohtana. Käsitteen jälkeen selostan *termin* käsitettä varsin laajasti, koska termi on keskeinen käsite tutkimukseni terminpoimintaa käsittelevässä osassa. Käsitesuhteita ja käsitepiirteitä tarkastelen koettimien käsittelyn yhteydessä jaksoissa 7.1.1 ja 7.1.2.

## 2.1 Käsitteen käsite terminologian teoriassa

Filosofit ovat pohtineet käsitteen olemusta jo ainakin 2500 vuotta. Heidän lisäksi *käsitteen* käsitteen olemus on askarruttanut muun muassa psykologeja, loogikoita, neurologeja ja terminologeja, mutta käsitteen perimmäisestä olemuksesta ei ole päästy yksimielisyyteen, sillä esimerkiksi Pilke (2000: 54) on löytänyt käsitteelle peräti 18 määritelmää yksistään terminologian kirjallisuudesta.

Terminologisessa kirjallisuudessa on tapana puhua käsitteestä tutkimuskohteena ja sanastotyön lähtökohtana, vaikka yleensä terminologit ovat kiinnostuneita vain yleiskäsitteistä (general concepts), eivät yksilökäsitteistä (individual concepts), joiden alaan kuuluu vain yksi tarkoite (ISO 1087-1:2000(E/F): 2–3). Yksilökäsite eroaa yleiskäsitteessä myös siten, että yksilökäsite voidaan sijoittaa aikaan ja paikkaan, mutta yleiskäsite on periaatteessa ajasta ja paikasta riippumaton (Pearson 1998: 129). Kiinnostukseni kohteena ovat tässä tutkimuksessa vain yleiskäsitteet, vaikka viittaan niihin sanalla *käsite*, niin kuin terminologisessa kirjallisuudessa yleensäkin.

Käsitteitä voidaan jaotella luokkiin käyttöalan mukaan, jolloin erotetaan yleistieteelliset, monen alan yhteiset ja tietylle alalle ominaiset käsitteet (ks. esim. Lejčik & Šelov 1990: 23). Sanastotyössä keskitytään yleensä alan ominaiskäsitteisiin, mutta myös yhteiskäsitteitä otetaan sanastoon. Tällöin on muistettava, että käsitesisältö voi muuttua, kun käsite siirtyy alalta toiselle. Tällä jaottelulla on merkitystä terminpoiminnassa, jossa ei aina ole yksinkertaista päättää, mitkä termit nimeävät tietyn alan ominaiskäsitteitä ja mitkä vähintään kahdelle alalle yhteisiä käsitteitä.

Käsitteet on perinteisesti jaettu neljään luokkaan sen mukaan, ovatko niihin liittyvät reaali maailman tarkoitteet olioita (entity; объект), prosesseja, toimintaa, tapahtumia (activity; процесс), ominaisuuksia (property; признак) vai suureita (dimension; величина) (ks. esim. Lejčik & Šelov 1990: 33). Suureiden luokkaa ei aina tarvita, sillä muun muassa Griněv (2002: 20), Montero-Martínez, García de Quesada ja Fuertes-Olivera (2002: 187) sekä Karihalme (1996: 27) luokittelevat tarkoitteet kolmeen ontologiseen luokkaan eli olioihin (объекты; objects), prosesseihin, tapahtumiin tai toimintoihin (процессы; events) ja ominaisuuksiin (свойства; properties). Suurinta osaa käsitteistä vastaa oliotarkoite. Citkina (1988: 35–37) esittää Kandelakin (1970: 37) matemaattista logiikkaa käsitelleen tutkimuksen pohjalta taulukon, jonka mukaan noin 85 prosenttia tarkoitteista on olioita, viisi prosenttia prosesseja, kuusi prosenttia ominaisuuksia, kolme prosenttia suureita ja kaksi prosenttia tieteitä tai erikoisaloja. Näin ollen ei ole mikään ihme, että oliot ovat saaneet osakseen huomattavasti enemmän terminologien huomiota kuin muut ontologiset luokat.

Tässä tutkimuksessa käsitteiden luokittelulla tarkoitteita vastaaviin luokkiin on merkitystä terminologisia koettimia käsittelevässä osassa, sillä käsitepiirteet ja käsitteiden väliset suhteet riippuvat käsiteluoosta. Tutkimukseni terminologisia koettimia käsittelevässä luvussa noudatan käsitteiden jaottelussa perinteistä tarkoitteiden mukaista ryhmittelyä, jonka mukaan tarkoitteet ovat olioita, toimintoja, tapahtumia, ominaisuuksia tai joskus harvoin suureita.

Kuten uusi terminologinen tutkimus osoittaa, monen termien ja käsitetiedon poimintaan liittyvän ongelman perimmäinen syy on käsitteiden tasolla. Muun muassa käsittei-



den muuttuminen ja näkökulmaerot vaikeuttavat termien ja käsitetiedon poiminnan automatisointia. Erityisesti käsitetiedon poiminnassa käsitteiden olemuksen tunteminen on tärkeää, sillä erilaisista käsitteistä annetaan teksteissä erilaista tietoa.

Seuraavassa jaksossa selostan terminologian tutkijoiden käsityksiä käsitteen olemuksesta termien ja käsitetiedon poiminnan näkökulmasta. Aloitan klassiseen luokitteluun perustuvan perinteisen terminologian teorian mukaisesta näkemyksestä. Tämän jälkeen selostan viime vuosikymmeninä esitettyjä näkemyksiä, jotka laajentavat perinteisen terminologian teorian mukaista *käsitteen* käsitettä.

## 2.2 Klassisesta luokittelusta elinkaariajatteluun

Terminologien käsitys käsitteistä on perustunut klassiseen luokitteluun, jonka lähtökohdiana puolestaan pidetään yleisesti Aristoteleen ajattelua. Aristoteles esitti, että ajattelu perustuu assosiaatioihin, joita ovat erityisesti yhteenkuuluvuuden ja samankaltaisuuden periaatteet. Klassista luokittelua (classical theory of classification) on käsitelty useissa teoksissa (ks. esim. Lakoff 1987: 157; myös Bowker 1996a: 781–782). Teorian pääkohdina pidetään näissä lähteissä seuraavia periaatteita:

- Käsitteet määritellään välttämättömien (necessary) ja riittävien (sufficient) piirteiden avulla.
- Käsitteellä joko on tietty käsitepiirre tai sitä ei ole.
- Käsitteillä on selvät ja kiinteät rajat.
- Kaikki saman luokan jäsenet ovat samanarvoisia.
- Käsitteet ovat olemassa ihmisestä riippumatta.
- Käsitteillä on yksi oikea luokittelu.

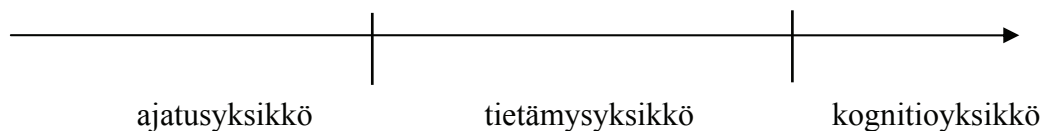
Klassinen luokitteluteoria oli vallitseva 1950-luvulle saakka ja terminologiassa jopa pitempään. Kun terminologien kiinnostus laajeni tekniikan alojen ulkopuolelle, teoria osoittautui jäykäksi eikä se toiminut kaikissa tilanteissa, sillä osa käsitteistä ei sopinut perinteiseen käsitteen malliin. Edellä esitettyjä periaatteita ovat kritisoineet muun muassa Wittgenstein (1953) (perheyhtäläisyys ja peliteoria), Rosch (1978) (prototyypiteoria), Lakoff (1987) (kognitiiviset mallit) ja Barsalou (1983). Joitakin käsitteitä ei voinutkaan määritellä välttämättömien ja riittävien piirteiden avulla. Käsitepiirteiden kaksinapaisuuskaan ei ole aivan niin yksioikoista, kuin näyttää. Esimerkiksi Barsaloun (1983: 213) mukaan käsitteillä voi olla kontekstista riippumattomia tai kontekstista riippuvia piirteitä ja vain osa niistä on kerrallaan aktiivisia. Käsiteluokat voivat olla joko-tai-luokkia (tarkoite joko kuuluu luokkaan tai ei) tai liukuvarajaisia (graded), jolloin käsiteluokan rajat voivat olla selkeät tai sumeat (vrt. Koivisto-Alanko 2000: 58). Klassisen teorian mukaan on vain yksi ”oikea” tapa luokitella, mutta käytännössä ihmiset luokittelevat käsitteitä hyvin eri tavoin näkökulmasta riippuen. Yhdessä luokittelussa pitäytymiseen ei olekaan mitään syytä, sillä kuten Lakoff (1987: 287) toteaa, klassiset luokittelut ja taksonomiat ovat ihmismielen keksintöä.

Käsiteluokittelun ongelmia on yritetty ratkaista ottamalla käyttöön luokkia, joiden alaan sopivat sellaisetkin käsitteet, jotka eivät sovi perinteisen käsitteen alaan. *Käsite* (concept) voi olla yläkäsite, jonka alakäsitteitä ovat *ajatusyksikkö* (unit of thought), *tietämysyksikkö* (unit of knowledge) ja *kognitioyksikkö* (unit of cognition), kuten Picht (2002) ehdottaa. Temmermanilla (2000: 43, 65) yläkäsite on *ymmärrysyksikkö* (unit of understanding), jonka alakäsitteitä ovat *käsite* (concept; perinteisen terminologian mukainen ymmärrysyksikkö) ja *käsiteluokka* (category; prototyypirakenteinen ymmärrysyksikkö). Temmermanin luokittelun mukainen käsite voidaan määritellä perinteisesti sisällön ja alan avulla. Käsiteluokka sen sijaan rakentuu prototyypin ytimen ympärille. Temmerman soveltaa luokitteluunsa prototyypiteoriaa, jonka mukaan osa tarkoiti-

teista on luokan keskellä ja osa reunoilla. Käsite on sitä keskemällä käsiteluokassa mitä enemmän sillä on yhteisiä käsitepiirteitä luokan prototyyppikäsitteen kanssa (Mannerko 2000: 38). Prototyyppikin voi vaihdella eri aikoina tai eri kulttuureissa, sillä ihmisten käsitys maailmasta muuttuu ja vaihtelee kulttuurista riippuen.

Pichtin luokittelu perustuu käsitteiden elinkaareen, jonka ensimmäinen vaihe on ehkä vain yhden ihmisen ajattelun synnyttämä ajatusyksikkö, toinen vaihe on perinteisen käsitteen tavoin erikoisalalla käytettävä tietämysyksikkö ja kolmas vaihe on muutokseen kypsynyt kognitioyksikkö. (Picht 2002: 4–10). Kaikki käsitteet ovat Pichtin luokittelun mukaan aluksi ajatusyksiköitä, koska käsite syntyy ensin yhden asiantuntijan ajatusprosessin tuloksena ja saa ehkä pian myös sanallisen määritelmän. Ajatusyksikkö ei voi Pichtin mukaan olla sanastotyön lähtökohta subjektiivisuutensa vuoksi. (Picht 2002: 3–4.) Kaikki varsinaiset käsitteet perinteisessä mielessä ovat tämän luokittelun mukaan tietämysyksiköitä, joiden määritelmät ovat erikoisalan toimijoiden hyväksymiä, vaikka määritelmien muotoilu ei välttämättä ole tapahtunutkaan kollektiivisesti. Tietämysyksikön sisältö on kaikki se tietämys, mikä ihmisillä tiettyä aikana tästä tietämysyksiköstä on maailmanlaajuisesti, vaikka kaikki tieto ei kaikissa viestintätilanteissa olekaan aktiivista. Kognitioyksikön elinkaari on käsitteen elinkaarta pitempi ajanjakso. Siinä otetaan huomioon myös käsitteen muuttuminen tiedon lisääntyessä sekä se, ettei käsite ole vain ajatusmaailman irrallinen yksikkö, vaan se reaalistuu aina, kun siitä viestitään jossakin viestintätilanteessa. Tällöin käsite on koko ajan alttiina muutoksille. Pichtin mukaan kognitioyksikkö on jo saavuttanut elinkaarensa pään ja muuttuu uudeksi käsitteeksi. (Picht 2002: 8–10.)

Myös Lejčikilla ja Šelovilla (1990: 33) on Pichtin luokittelua vastaava luokittelu, jota he nimittävät sosiologiseksi luokitteluksi. Niin ikään Ozeki (1987 lähteessä Pilke 2000: 50) on samoilla linjoilla Pichtin kanssa, kun hän sanoo, että käsitteellä on aina kolme ulottuvuutta: yksilöllinen, kansallinen (sidottu äidinkieleen) ja inhimillinen. Hänen mukaansa käsite on sekä ajattelun että tiedon yksikkö, joka sisältää kaiken senhetkisen tiedon käsitteestä. (Ks. myös Ivina 2003: 41.)



KUVIO 3. Käsiteluokkien sijainti aikajanalla elinkaariajattelun mukaan

Pichtin malliin sisältyy implisiittisesti myös sellainen käsitteen ominaisuus, jota siihen ei perinteisesti ole totuttu liittämään: käsitteet eivät pelkästään synny ja kuole, vaan tämän lisäksi ne voivat muuttua. Tämän ajatuksen kanssa Picht ei ole yksin, sillä esimerkiksi Lotte (1982: 17) on sanonut, että käsite voi muuttua joko vähitellen tai yhtäkkiä. Samoilla linjoilla ovat Kvitko, Lejčik ja Kabancev (1986: 20), joiden mukaan vaatimus termin sisällön pysyvyydestä tarkoittaa termin merkityksen pysyvyyttä tietyssä käsitejärjestelmässä ja erikoisalan kehityskaaren tietyssä vaiheessa. Käsitteiden muuttumattomuus on siis myytti, joka syntyi, koska perinteinen terminologia keskittyi tarkasteluhetkeen ja kielenohjailun avulla tulevaisuuteen, josta ei voi tehdä havaintoja.

Käsitteiden muuttumista ei juuri ole tutkittu ennen viime vuosikymmentä. Perinteisesti tutkimuksen lähtökohta on ollut, että käsite ei muutu eikä terminkään tarvitse muuttua, vaan hyvä termi on käyttökelpoinen määräämättömän ajan. Termivariaation

terminpoiminnalle aiheuttamat ongelmat on pyritty ratkaisemaan standardisoinnin keinoin opastamalla kielenkäyttäjiä käyttämään vain tiettyjä terminuotoja ja hylkäämään kilpailevat muodot. Samoin synonymiaa on pidetty ei-toivottuna ilmiönä, koska mahdollisia erilaisia näkökulmia ja niiden mukaisia erilaisia luokitteluja ei ole otettu huomioon. Uuden käsitteistöityksen mukaan termivariaatio ja termien synonymia ovat luonnollisia ilmiöitä erikoiskielessä (ks. esim. Jacquemin 2001: 313; Temmerman 2000: 133). Kun ne pystytään selittämään käsitteiden muutoksesta ja näkökulmaeroista johtuviksi, niiden terminpoiminnalle aiheuttamat ongelmat voidaan ehkä ratkaista muutenkin kuin kielenohjailun avulla. Nykyisin kehitys on monella alalla niin nopeaa, että käsitteet saattavat muuttua jo muutaman vuoden aikana, minkä vuoksi termistöäkin on pakko korjata ja järjestää uudelleen.

Pichtin ja Temmermanin lisäksi käsitteiden luokittelua ovat kehittäneet muun muassa Lakoff (1987) ja Manerko (2000). Tutkijoiden ajatukset on otettu huomioon uusimmissa standardeissakin. Esimerkiksi kansainvälisessä standardissa (ISO 1087-1: 2) käsite määritellään tietyn, vain tälle käsitteelle kuuluvan piirrejoukon muodostamaksi tietämysyksiköksi.

Pichtin elinkaariajatteluun ja Temmermanin prototyypiteoriaan perustuva käsitteeluokittelu on mielenkiintoinen ja avaa uusia näkökulmia terminologiseen tutkimukseen. Käsitteiden elinkaaren tai prototyypisyyden tarkastelu jää kuitenkin tämän tutkimuksen ulkopuolelle. Eräiden merenkulun turvallisuuskäsitteiden nimitysten muutosta elinkaaren eri vaiheissa olen selostanut artikkelissani *Changing Terms in the Finnish VTS Terminology* (Pasanen 2003).

Prototyypiteoriaa soveltavat terminologit, kuten esimerkiksi Temmerman (2000), lähestyvät psykolingvistiikan edustajia (esimerkiksi Rosch), jotka luokittelevat käsitteitä abstraktiotason mukaan. Heidän luokittelussaan kaikkein alhaisin abstraktiotaso on erisnimien nimeämällä yksilökäsitteillä. Terminologia on perinteisesti rajoittunut abstraktiotason keskivaiheille eli käsitteisiin, joilla on selkeät tarkoitteet mutta jotka eivät ole yksilökäsitteitä.

Käsitteiden abstraktiotasoja on tutkittu varsinkin yleiskielessä, mutta periaate on sovellettavissa myös erikoiskieliin. Manerkon (2000: 41) mukaan Rosch ja muutama muu tutkija pitävät perustason käsitteiden nimityksiä prototyypikategorioiden perustana. Perustason käsitteiden nimitykset kuuluvat yleiskieleen ja elämänkokemukseen, ne ovat lyhyitä, helppoja muistaa ja niiden käsittely vaatii vain vähän ajattelutyötä. Lakoffin (1987: 46, 270–271) mukaan perustason käsitelukat (basic-level categories) on olioiden (objects) lisäksi toiminnoilla (actions) ja ominaisuuksilla (properties). Esimerkiksi liikkuminen on ylätasoinen käsite, kun taas juokseminen ja käveleminen ovat perustason käsitteitä. Erilaiset tavat kävellä ovat perustason käsitteiden alakäsitteitä.

Jos yleiskielestä havaitut abstraktitasot pätevät erikoiskieliin, niin perustason käsitteiden nimitykset ovat lyhyitä, käyttöön vakiintuneita ja suurifrekvenssisia, mutta yleiskieleen kuuluvina ne eivät välttämättä ole termejä. Maallikon käsitemaailma on perustasolla, ja tarkempi luokittelu on tarpeen vain asiantuntijoille. Maallikolle esimerkiksi kaikki lastia kuljettavat alukset ovat *laivoja*, kun taas merenkulun asiantuntijalle tämä ei riitä, vaan lastauksen suunnittelun kannalta on tärkeä tietää, onko alus *ro-ro-alus* vai *konventionaalinen kuivarahtialus*. Saman oletuksen mukaan käsitteiden nimitykset ovat sitä pitempiä ja pienifrekvenssisempiä mitä alemmas käsittehierarkiassa edetään. Oletettavasti myös variaation yleisyys riippuu abstraktiotasosta, sillä pitemmät nimitykset antavat enemmän variaatiovaihtoehtoja kuin lyhyet.

Olen edellä kuvaillut lyhyesti erilaisia terminologisessa kirjallisuudessa esitettyjä käsiteluoitteluja. Elinkaaren vaihe, abstraktiotaso ja prototyypisyys ovat luokitteluperusteita, jotka on otettu käyttöön, koska klassinen käsitteiden luokittelu on osoittautunut

riittämättömäksi. Nykyisin onkin vahvistunut näkemys, että käsitteillä ei ole yhtä oikeaa kaikissa tilanteissa pätevää luokittelua. Sen sijaan käsitteitä luokiteltaessa niiden ajatellaan kuuluvan tietylle erikoisalalle, niitä tarkastellaan tietyistä näkökulmista ja niillä on tietty kiinteä piirteiden joukko. Tästä syystä miltei kaikki luokittelut ovat kompromisseja, sanoo Hirs (1993: 225), joka on tutkinut lääketieteen erikoiskieltä:

”Objects have to be considered within a specific field, a given point of view and a fixed sequence of subdividing characteristics. This is why the ICD (International Classification of Diseases), like (almost) every classification, is a compromise between conflicting interests.” (Hirs 1993: 225)

Käsite voidaan aina luokitella tarpeen mukaan uudelleen riippumatta siitä, miten se on aiemmin luokiteltu (Barsalou 1983: 226). Esimerkiksi alukset voidaan luokitella hyvin monella tavalla tarkoituksesta riippuen, eikä kaikkia luokitteluja ole varmasti vielä tehty. Tarkoitus määrää, mikä luokittelupiirre kulloinkin on määräävä, onko se lastinantajalle tärkeä lastityyppi vai ympäristöviranomaisille tärkeä luokittelu mahdollisen ympäristöonnettomuuden varalta. Näkökulma vaikuttaa myös määrittelyyn, jolloin käsitteellä voi teksteissä olla monta määritelmää ja käsitejärjestelmässä käsitteet voidaan järjestää eri piirrelajin mukaan. Erilaisissa luokitteluissa yläkäsitekin saattaa vaihtua, mutta käsite pysyy silti samana.

Uusien käsiteluokittelujen ansiosta on mahdollista soveltaa elinkaariajattelua, jota esimerkiksi Picht (2002) on tehnyt tunnetuksi. Elinkaariajattelun avulla voidaan selittää monia termien ja käsitetiedon poimintaan liittyviä ongelmia. Käsitteiden syntymän, elämän ja kuoleman kysymysten tunteminen auttaa ymmärtämään termivariaatiota ja synonymiaa, jotka vaikuttavat terminpoimintaan. Käsitteen elinkaaren vaihe vaikuttaa myös siihen, mitä käsitetietoa käsitteestä annetaan teksteissä. Näihin ilmiöihin palaan työni empiirisessä osassa termien ja käsitetiedon poiminnan yhteydessä luvuissa 6 ja 7. Tässä luvussa jatkan terminologian teoreettisten kysymysten pohdintaa ja esittelen seuraavaksi *termin* käsitteen terminologian teorian näkökulmasta.

## 2.3 Käsite *termi* terminologian teoriassa

Olen edellä kuvannut terminologian peruskäsitteistä *käsitteen* käsitettä terminologisen tiedon poiminnan näkökulmasta. Seuraavaksi tarkastelen *termin* käsitettä terminologiasa. Termi saa runsaasti huomiota jo työni tässä vaiheessa, sillä sen painoarvo työni empiirisessä osassa on erittäin suuri. Selostan ensin lyhyesti, miten termi on määritelty kansainvälisissä standardeissa, millainen on tutkimuskirjallisuuden mukainen ihanne-termi ja mikä on kognitiivisen terminologian näkökulma termiin.

### 2.3.1 Standardinmukainen näkemys termistä

Termin olemusta on pohdittu luultavasti kaikissa terminologisissa tutkimuksissa, mutta kukaan ei ole pystynyt antamaan määritelmää, jonka avulla mistä tahansa sanasta tai sanaliitosta voisi heti varmasti sanoa, onko se termi vai ei. Yksimielisiä ollaan vain siitä, että termi nimeää erikoisan yleiskäsitteen.

Määrittelemisen vaikeutta kuvaavat kansainvälisen standardisoiemisjärjestön ISO:n määritelmät. Vuonna 1999 hyväksytyn standardin ISO 12620 mukaan termi on määritellyn käsitteen nimitys (ISO 12620:1999(E): 5). Tämän määritelmän mukaan käsitteen nimitys on termi vasta, kun käsite on määritelty. Kaikkia erikoisan käsitteitä ei kuitenkaan ole määritelty tarkasti ja yksiselitteisesti, vaan joidenkin käsitteiden rajat ovat hämäriä, tai niiden käsitesisällöstä ollaan erimielisiä. Itse asiassa monet erikoisan kes-

keisistä käsitteistä eivät täytä tätä vaatimusta. (Ks. esim. Andreev 1999: 125–126.) Esimerkiksi *hyvä merimiestapa* on merenkulussa keskeinen käsite, koska se esiintyy hyvin monessa merenkulun turvallisuutta käsittelevässä tekstissä, mutta tarkkaa määritelmää sille on hyvin vaikea antaa.

Vain vuotta myöhemmin hyväksytyssä standardissa ISO 1087-1 termi on määritelty uudelleen. Uudemman määritelmän mukaan termi on jollakin erikoisalalla yleiskäsitteen nimityksenä käytettävä kielellinen ilmaus (ISO 1087-1:2000(E/F): 6). Määritelmässä on neljä ehtoa, joiden kaikkien on täyttyttävä, jotta sana tai sanaliitto voi saada termin statuksen. Ensinnäkin määritelmä sitoo termin johonkin erikoisalaan. Toisin sanoen erikoisalojen ulkopuolella ei voi olla termejä. Toiseksi määritelmän mukaan termit ovat vain yleiskäsitteen nimitykset. Yksilökäsitteiden nimitykset, esimerkiksi järjestöjen nimet, jäävät vaille termin statusta. Kolmanneksi termi on määritelmän mukaan aina kielellinen ilmaus, ei esimerkiksi piirrossymboli. Neljänneksi termin on oltava käytössä. Termin statusta ei ansaitse esimerkiksi sanakirjaan laitettu vastine, jota ei erikoisalan viestinnässä käytetä. Standardin määritelmä on kuitenkin varsin yleisluontoinen eikä riitä ratkaisemaan monia termintunnistuksen ongelmatilanteita. Määritelmä sulkee pois yksilökäsitteiden nimitykset, mutta toisaalta termiyn saavuttamiseksi riittää, että ilmaisua käytetään erikoisalalla yleiskäsitteen nimityksenä riippumatta ilmaisun kielellisestä muodosta, käytön yleisyydestä tai määritelmän olemassaolosta.

Jos termintunnistuksen ainoaksi ohjenuoraksi otetaan määritelmän erottava piirre, jonka mukaan termi nimeää erikoisalan yleiskäsitteen, terminpoimijalta vaaditaan hyvää erikoisalan tuntemusta. Aiemmat terminpoimintakokeet ovat kuitenkin osoittaneet, ettei termintunnistus ole helppoa edes erikoisalan asiantuntijalle (ks. jakso 6.3).

Perinteisen käsityksen mukaan termit ovat samanarvoisia, mutta eivät samanlaisia. Standardin ISO 12620 mukaan termit jakautuvat erilaisiin termityyppeihin, joita ovat päätermi (head term), synonyymi, kvasisynonyymi, kansainvälinen tieteellinen termi, yleisnimi ja internationalismi. Kvasisynonyymi edustaa samaa tai lähes samaa käsitettä kuin toinen samankielinen termi, mutta termien korvattavuus toisillaan rajoittuu joihinkin konteksteihin. Yleisnimi on tieteellisen termin synonyymi yleiskielessä. (ISO 12620: 1999(E): 5–6.) Lisäksi termillä voi pitkän muodon ohella olla käytössä lyhennetty muoto (abbreviated form of term). Samoin kuin käsitteet myös termit voivat olla alapidonnaisia tai monen alan yhteisiä (Pearson 1998: 25). Jotkut termit ovat selkeästi tietyn erikoisalan termejä (päätermejä tai synonyymeja) ja loput ovat kyllä termejä, mutta niihin liittyy jokin ominaisuus, joka tekee niistä ongelmallisia. Tällaisia ovat esimerkiksi monen alan yhteiset tai kansainväliset tieteen termit sekä kansainväliset lainat, mutta myös yleiskielen sanat termeinä tai termin osina.

Näiden varsin laveiden termin määritelmien perusteella on ymmärrettävää muun muassa Pearsonin (1998: 21) ärtymys. Hän arvostelee tutkijoita siitä, että käsite *termi* oletetaan tunnetuksi ja keskustelu on keskittynyt enemmän termien luokitteluun kuin termin määrittelyyn. Termien luokittelulla on varmasti paikkansa terminologiassa, mutta terminpoiminnassa luokittelusta ei ole apua, sillä termit on tunnistettava ennen kuin ne voidaan luokitella. En ole tutkimuksessani keskittynyt termien tai termiehdokkaiden luokitteluun, koska siitä ei näkemykseni mukaan ole suoraan apua terminpoiminnassa. Standardissa mainituista termityypeistä tarkastelen lähemmin vain lyhenteitä ja synonyymeja.

Olen edellä puhunut yleisesti termistä ja tarkoittanut sillä ISO-standardin 1087-1 määritelmän mukaisesti ilmausta, joka nimeää erikoisalan yleiskäsitteen. Terminologiassa termi on siis määritelty eri tavalla kuin tiedonhaussa käytetyt indeksitermit, jotka on määritelty tietyn tekstin sisältöä kuvaaviksi ilmauksiksi. Niiden tarkoitus on ohjata käyttäjä tiedon lähteille asiakirjan sisällä. (Lahtinen 2000: 7.) Terminologiassa termi



nimeää käsitettä kontekstista riippumatta, kun taas indeksitermit kuvaavat ensi sijassa tietyn asiakirjan sisältöä. Termit ovat usein myös kelvollisia indeksitermeiksi, mutta kuten Lahtinen (2000: 34) sanoo, termit ovat vain osa indeksitermeistä ja toisaalta on myös termejä, jotka eivät ole indeksitermejä. Selvyyden vuoksi tiedonhaussa puhutaankin indeksitermien ohella teknisistä termeistä, jotka ovat terminologian teorian määritelmän mukaisia termejä. Tässä tutkimuksessa tarkoitan termillä nimenomaan teknisiä termejä.

### **2.3.2 Ihannetermi sanastotyön oppaiden ja tutkimuskirjallisuuden mukaan**

Kansainvälisten standardien termille antamat määritelmät eivät juuri helpota terminologin tai kääntäjän työtä silloin, kun tehtävänä on tunnistaa termit. Terminpoiminnan ongelmien kanssa kamppaileva terminologi, kääntäjä, tulkki, tekninen kirjoittaja tai muu ohjeita terminpoimintaan kaipaava saattaa standardien lisäksi turvautua sanastotyön oppaisiin tai tutkimuskirjallisuuteen toivoen, että niissä kerrottaisiin, miten termit voi tunnistaa. Terminologian teoriassa käsitys termistä on 1980-luvulle saakka liittynyt perinteisen terminologian teoriaan ja sen mukaiseen näkemykseen käsitteen, termin ja tarkoitteen suhteista. Termille on luotu ihanne, jota esitellään sanastotyön oppaissa (esim. Haarala 1981: 11) ja tutkimuskirjallisuudessa (esim. Karihalme 1996: 336). Sagerin (1990: 89–90) mukaan ihannetermi

- nimeää käsitteen selkeästi ja kuvaavasti
- on leksikaalisesti systemaattinen eli termijärjestelmä noudattaa käsitejärjestelmää
- noudattaa kielen sananmuodostussääntöjä
- on sopiva johdosten muodostamiseen
- ei sisällä toistoa
- on lyhyt
- on yksimerkityksinen eli yksi termi nimeää vain yhtä käsitettä
- on tarkkarajainen
- ei riipu kontekstista.

Ihannetermillä ei myöskään ole morfologisia variantteja tai synonyymeja eli yhtä käsitettä nimeää vain yksi termi.

Tässä luettelossa kiteytyvät perinteisen terminologian teorian termivaatimukset, joihin jotkut tutkijat haluavat lisätä vielä määritelmän olemassaolon. Luetellut vaatimukset ovat kuitenkin nimenomaan hyvän termin ihanteellisia ominaisuuksia, eivät käytössä olevien termien todellisia tunnuspiirteitä. Sager (1990: 89) myöntääkin, että ihannetilanne on mahdollinen vain erityistapauksessa. Monen muun tutkijan tavoin Šelov (2001: 2) luettelee hyvän termin ominaisuudet, mutta pitää niitä aivan liian yleisluontoisina, jotta niitä voitaisiin käyttää termintunnistuksessa. Grinëv (1993: 34) varoittaakin sekoittamasta todellisten termien piirteitä ihannetermille asetettujen vaatimusten kanssa, niin kuin joissakin töissä on tehty. Toivottuja ominaisuuksia voidaan toki luetella, mutta termiä ei voi sanoa kelvottomaksi tai tarpeettomaksi, ellei sillä ole näitä ominaisuuksia. Hyvän termin vaatimusten avulla ei termejä pysty tunnistamaan tekstistä, koska ne eivät ole mitattavia. Sen sijaan ne ovat suhteellisia, kielestä riippuvaisia, subjektiivisia ja keskenään ristiriitaisiakin (kuten lyhyys ja tarkkuus).

Perinteisen terminologian teorian mukaisen terminäkemyksen ongelma on Pearsonin (1998: 12–16) mukaan se, että siinä käsitteiden sisällöstä ja niiden välisistä suhteista sovitaan ennen nimeämistä. Termi on nimilappu, joka annetaan käsitteelle. Ihannetapauksessa käsitejärjestelmä on tämän jälkeen hierarkkisesti järjestynyt ja sitä vastaa leksikaalisesti systemaattinen termijärjestelmä. Nämä kaksi järjestelmää sopivat hyvin päällekkäin, koska yhtä käsitettä vastaa yksi termi ja yksi termi nimeää yhden käsitteen.

Kun käsitejärjestelmä on vielä kansainvälisesti harmonisoitu, sen päälle voidaan asettaa minkä tahansa kielen mukainen termijärjestelmä tarpeen mukaan. Käsitejärjestelmä ja termijärjestelmä pysyvät muuttumattomina, eikä viestinnässä tapahdu väärinkäsityksiä. Eri alojen sanastostandardit ovat askeleita kohti tätä ihannetilaa, mutta kuten Pearson (1998: 12) toteaa, standardeihin liittyy ongelma: käytännön viestinnässä kieltä ei ainaakaan toistaiseksi käytetä niiden mukaisesti.

Jos termejä tarkastellaan ihannetermin näkökulmasta, termeiksi kelpaa vain osa käsitteiden nimityksistä. Näkökulma rajoittaa termin käsitteen suppeammaksi, kuin mitä se on todellisessa kielenkäytössä. Vakiintuneiden ja hyvin muodostuneiden termien ohella todellisessa kielenkäytössä käsitteitä nimetään usein elliptisillä ilmauksilla, joista puuttuu jokin termin osa. Ihannetermillä ei ole synonyymeja mutta todellisuudessa synonyymia ja termivariaatio on tavallista (ks. esim. Sewangi 2001: 95). Myös termin variantit nimeävät samaa käsitettä, mutta leksikaalisesti ne poikkeavat toisistaan vain vähän, kun taas synonyymiset termit voivat poiketa paljonkin. Ihannetilanteen vastaisesti todellisessa kielenkäytössä esiintyy polysemiaa, jossa yksi termi nimeää vähintään kahden toisiinsa semanttisesti liittyvää käsitettä. Yhden erikoisalan sisällä polysemia ilmenee muun muassa siten, että sama termi voi nimetä sekä toimintaa että toiminnan tulosta (L'Homme 2003: 405). Termejä myös käytetään ”väärin”. Kirjoittaja ei ehkä muista oikeaa termiä tai haluaa tietoisesti käyttää eri termiä, kuin jostakin erikoisalan kysymyksestä mahdollisesti eri mieltä oleva toinen alan asiantuntija. Näistä syistä jotkut tutkijat, kuten esimerkiksi Feliu ja Cabré (2002), puhuvatkin terminologisista yksiköistä (terminological units) termien sijaan (vrt. Grinev 2004: 53).

Kvitko ym. (1986: 18–38) ovat valmiit joustamaan monista ihannetermin vaatimuksista, jotka ovat kaukana erikoisalojen todellisesta kielenkäytöstä. Ensinnäkään he eivät esitä termeille ja termijärjestelmille kielellisen systemaattisuuden vaatimusta, mutta loogisuuden vaatimus on heidänkin mielestään tärkeä. Toiseksi johdosten muodostaminen koskee heidän mielestään kaikkia sanoja eikä ole mitenkään vain termien ominaisuus. Sama koskee ääntämisen helppoutta ja kielellistä moitteettomuutta, vaikka he asettavatkin erikoiskielen normit yleiskielen normien edelle. Heidän mukaansa on myös paljon termejä, jotka eivät sovi johdosten muodostamiseen. Termi voi esimerkiksi olla sanaliitto eikä siksi sovi johdosten muodostamiseen. Kolmanneksi lyhyys toteutuu Kvitkon ym. mukaan rajoitetusti tai ei ollenkaan. Heidän mukaansa optimaalinen pituus on kyllä tavoite, mutta lyhyttä tärkeämpää on esimerkiksi motivoituneisuus eli se, että termi kuvaa hyvin käsitettä. He ovat valmiita tinkimään myös yksimerkitysisyyden vaatimuksesta ja sanovat, että eri käsitejärjestelmiin voi kuulua samoja termejä, koska jokaisessa käsitejärjestelmässä käsitteellä on eri määritelmä, mutta sekaannusta ei saa syntyä.

Ihannetermin vaatimukset ovat sinällään hyviä ja toivottavia, mutta luettelo luo kuvaa ideaalitalanteesta, ja moni tutkija on havainnut, että edellä esitetyt perinteisen terminologian mukaiset näkemykset termistä eivät todellisessa erikoisalojen kielenkäytössä päde. Soininen (1998: 3) sanookin, että terminmuodostussäännöillä on hyvin vähän tekemistä sen kanssa, miten tai millaisessa ympäristössä termiä käytetään. Esimerkiksi Sewangi (2001: 94–96) joutui toteamaan, että hänen aineistostaan poimitut termit eivät täyttäneet useimpia perinteisen terminologian mukaisen ihannetermin vaatimuksia. Sewangi havaitsi, että varsinkin synonyymia, variaatio, homonymia, polysemia ja kontekstisidonnaisuus ovat tavallisia ilmiöitä erikoiskielessä. Koska näiden ilmiöiden voidaan olettaa olevan alasta riippumattomia, niitä esiintyy todennäköisesti myös tämän tutkimuksen tutkimusaineistossa, joka on koottu merenkulun alalta.

Olen edellä selostanut sanastotyön oppaissa ja tutkimuskirjallisuudessa esitettyjä näkemyksiä termin olemuksesta ja todennut, että ne eivät juuri auta silloin, kun sanasto-

työn tekijän on ratkaistava, onko jokin sana tai sanaliitto termi vai ei. Terminpoiminnassa lähtökohta ei voi olla ihannetermi vaan erikoisalan käsitteen nimitys sellaisena, kuin sitä käytetään erikoisalaviestinnässä. Kognitiivinen näkemys termistä on lähempänä tätä lähtökohtaa kuin standardin tai sanastotyön oppaiden näkemys, sillä kognitiivinen näkemys sallii todellisessa kielenkäytössä esiintyvän vaihtelun. Järvi (2001: 86) kiteyttää toistenkin tutkijoiden havainnot sanomalla, että kognitiivisen terminologian ehkä tärkein idea on avara näkemys sanojen merkitysrajojen sumeudesta. Kognitiivisessa terminologiassa termejä ei enää ole perinteisen terminologian merkityksessä, koska sanoja käytetään aina viestintätilanteessa ymmärryksiköiden edustajina. Koska kontekstit vaihtelevat, myös sanojen merkitykset vaihtelevat ja rajat ovat enemmän tai vähemmän sumeat. (Ks. myös Temmerman 2000: 223–225.) Järvi (2001: 87) sanoo jopa, että terminologian ongelmat johtuvat paljolti ihmisen ajattelun sumeudesta, sillä ihmisen ajattelu ei sopeudu perinteisen terminologian mukaiseen malliin.

Olen tässä luvussa kuvannut tutkimukseni kannalta tärkeimpiä terminologian peruskäsitteitä, käsitettä ja termiä. Luvun alussa totesin, että käsitettä pidetään yleisesti terminologian lähtökohtana. Käsite on myös oman tutkimukseni lähtökohta, vaikka työn empiirisessä osassa keskityinkin termien ja käsitetiedon poimintaan. Ennen termien ja käsitetiedon poimintamenetelmien teoreettista pohdintaa tarkastelen kuitenkin vielä ikään kuin johdantona perinteisen ja uuden terminologian tavoitteita, tutkimuskohteita ja menetelmiä.



### 3 TERMINOLOGIAN TAVOITTEET, TUTKIMUSKOhteET JA MENETELMÄT

Terminologian juuret ovat kategorisaatiossa eli luokittelussa, joka on ihmisille tyypillistä ja auttaa hallitsemaan ympäristöä ärsykkeiden tulvassa. Tatarinov (1996: 266) toteaa, että luokitella voi kaikkea ja millä tahansa perusteella. Luokittelut ovat yleistäviä ja tapahtuvat osin tiedostamatta (Pälli 2003: 39). Tietoista luokittelua ja tiedon järjestämistä esimerkiksi erikoisalan sanakirjoihin on eri tieteenaloilla tehty vuosisatoja. Kuitenkin vasta monien tekniikan ja luonnontieteiden alojen nopea kehitys sekä teollinen vallankumous 1900-luvun alussa lisäsivät tiedon järjestämisen ja ennen kaikkea standardisoinnin tarvetta. Myös akateemisten piirien kiinnostus tiedon järjestämisen ja standardisoinnin ongelmiin kasvoi ja toimi lähtökohtana terminologian synnylle.

Terminologisen tutkimuksen ja sanastotyön kohteena ovat erikoisalojen käsitteet ja erikoiskielten termit. Erikoiskielet ovat terminologiassa keskeisessä asemassa, ja niiden tuntemus on tärkeää terminologisten tutkimusmenetelmien ja sanastotyömenetelmien kehittämisessä. Tästä syystä paneudun työssäni erikoiskielten ominaisuuksiin varsin laajasti. Terminologisen tutkimuksen ja sanastotyön menetelmät ovat muotoutuneet erikoiskielten piirteiden ja terminologian tavoitteiden mukaan ja muuttuneet tavoitteiden muuttuessa. Standardisointiin pyrkivän terminologisen sanastotyön rinnalle on tutkimuksessa ja myös sanastotyössä noussut uusia tavoitteita, jotka ovat vaatineet uusien menetelmien kehittämistä. Terminologisella tutkimuksella, samoin kuin sanastotyölläkin, voi tänä päivänä olla deskriptiivinen tavoite, jolloin pyrkimys on standardisoinnin sijasta kuvata tietyn erikoisalan käsitteistöä ja termistöä. Tutkimuksen ja sanastotyön kohteeksi on tekniikan alojen erikoiskielten lisäksi otettu esimerkiksi sosiaalialojen erikoiskieli (esim. Riggs 1993), jonka ominaisuudet eroavat tekniikan alojen erikoiskielten ominaisuuksista. Myös nämä uudet tutkimuksen ja sanastotyön kohteet ovat vaatineet menetelmien kehittämistä ja jopa teorian uudelleenarviointia.

On paradoksaalista, että alalla, jolla tutkimuskohteena ovat käsitteet sekä niiden nimitykset ja määritelmät, peruskäsitteiden sisällöstä ei ole päästy yksimielisyyteen. Jo tutkimusalan nimi *terminologia* on monimerkityksinen. Suomen kielessä *terminologia* tarkoittaa yleensä tutkimusala (Nykänen & Kalliokuusi 1999: 174). Sillä saatetaan kuitenkin viitata myös termien muodostamaan kokonaisuuteen, vaikka suomen kielessä tälle käsitteelle on oma nimitys: *termistö*. Englannissa *terminology* merkitsee tutkimusala, tietyn erikoisalan termistöä ja myös menetelmiä, joilla termejä kerätään ja kuvataan (de Bessé, Nkwenti-Azeh & Sager 1997: 154; Nykänen & Kalliokuusi 1999: 174). Venäjän kielessä termi *терминология* on aiemmin ollut kolmen käsitteen nimitys. Se on nimennyt sekä tietyn erikoisalan termijärjestelmää, koko kielen termistöä, siis esimerkiksi venäjän kielen kaikkia termejä yhdessä, että termejä tutkivaa tiedettä. Nykyisin kielen termistöstä käytetään venäjässä termiä *терминологика* ja tieteenalasta termiä *терминоведение*. (Grinëv 1993: 13.)

Seuraavaksi selostan terminologian tavoitteita, sen tutkimuskohteita ja käytössä olevia menetelmiä. Ensin luon lyhyen katsauksen perinteisen terminologian tavoitteisiin, tavoitteiden laajenemiseen ja nykyisiin terminologisen tutkimuksen ja sanastotyön monivaihteisiin tavoitteisiin. Toiseksi kuvaan terminologisen tutkimuksen ja sanastotyön tutkimuskohteiden, erikoisalojen käsitteiden ja erikoiskielten termien esiintymisympäristöä, erikoiskieliä. Niiden ominaisuuksien tunteminen auttaa ymmärtämään terminologisten tutkimusmenetelmien ja sanastotyömenetelmien kehittämisen ongelmakohtia. Kolmanneksi kuvaan terminologian menetelmiä, ensin perinteisen terminologian tarpeisiin kehitettyä käsitteanalyysia ja sitten tässä työssä käytettäviä termien ja käsitetiedon poimintamenetelmiä, joita kutsun uusiksi terminologisiksi menetelmiksi.

### 3.1 Perinteisen terminologian tavoitteet

Terminologian historian katsotaan alkaneen 1930-luvulla, jolloin itävaltalainen insinööri Eugen Wüster (1898–1977) esitteli terminologian periaatteet väitöskirjassaan *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik* (1931). Wüsterin periaatteet vastasivat tekniikan aloilla ilmenneisiin tiedon järjestämisen ja standardisoinnin tarpeisiin. Väitöskirja toimi alkusysäyksenä sen teorian synnylle, jota nykyisin kutsumme terminologiaksi, ja Wüsterin esittämien periaatteiden pohjalta syntyi ensimmäinen terminologian koulukunta, Wienin koulukunta.

Wüsterin periaatteet saivat kannatusta myös Neuvostoliitossa, kun hänen väitöskirjansa käännettiin venäjäksi vuonna 1935. Neuvostoliitossa oli jo aiemmin ollut kiinnostusta terminologiaan, mutta Wüsterin periaatteet lisäsivät aktiivisuutta niin, että Neuvostoliitossa syntyi Moskovan terminologinen koulukunta, jonka muotoutumiseen vaikutti erityisesti Dmitrij Semenovič Lotte. Kielimuurin vuoksi Moskovan koulukunnan tutkimuksia ei kuitenkaan länsimaissa juuri tunneta. Samanaikaisesti Moskovan koulukunnan kanssa syntyi silloisessa Tšekkoslovakiassa Prahan koulukunta. Tässä tutkimuksessa käytän näistä Wüsterin periaatteiden pohjalta syntyneistä terminologian koulukunnista Rita Temmermanin (2000) tavoin nimitystä *perinteinen terminologia* (traditional terminology).

Kaikki kolme mainitsemaani perinteisen terminologian koulukuntaa pyrkivät vaikuttamaan erikoisalojen kieleen kielenhuollon ja standardisoinnin kautta. Standardisointi nähtiin tarpeelliseksi erityisesti tekniikan aloilla, joilla kehitys oli muuttanut ihmisten toimintaympäristöä ja asettanut viestinnälle aivan uusia vaatimuksia. Uusilla aloilla tulivat pian tutuiksi monet viestintää hankaloittavat ilmiöt, joita ovat esimerkiksi Grinëvin (1993: 8–9) luettelemat homonymia, synonymia, koulukuntien väliset tulkintaerot, termivariaatio, puutteelliset määritelmät, vierasperäisten termien ujuttautuminen kieleen ja yleisten terminmuodostusperiaatteiden puuttuminen. Väärinymmärrysten välttämiseksi erikoisalojen viestinnälle oli luotava yhteisesti sovitut säännöt, ja terminologia vastasi tähän tarpeeseen. 1980-luvulle jatkunut tekniikan ja luonnontieteiden sekä standardisoinnin aseman korostuneisuus terminologian alalla juontaa siis historiasta.

Perinteisen terminologian koulukuntien edustajat uskoivat, että erikoisalojen viestintää voidaan parantaa juuri standardisoinnilla. Tavoitteena oli sopia yhteisesti, mitä nimityksiä erikoisalan yleiskäsitteistä käytetään, ja saada sitten jokainen erikoisalan viestintää harjoittava käyttämään näitä ja vain näitä nimityksiä eli termejä. Tavoitteena oli monosemia, jossa yhtä käsitettä nimetään vain yhdellä termillä, joka puolestaan nimeää vain yhtä käsitettä. Temmerman (2000: 4–17) luettelee standardisointiin keskittyneen terminologian tavoitteiksi erikoisalan sisäisen synonymian poistamisen, tarvittaessa uusien termien luomisen terminologian periaatteiden mukaan ja luodun sanaston saatamisen normin asemaan. Tutkittavien erikoisalojen luonteesta johtuen terminologinen tutkimus keskittyi aluksi aineellisten tarkoitteiden tutkimukseen.

Terminologian periaatteet ja menetelmät levisivät nopeasti Wienin Moskovan ja Prahan koulukuntien vaikutuspiiriin ulkopuolelle Euroopassa ja Kanadassa. Merkillepantavaa on, että periaatteet ovat saaneet eniten kannatusta toisaalta monikielisessä ja monen itsenäisen valtion Euroopassa ja toisaalta Neuvostoliitossa, jossa tavoitteena oli hallita laajaa valtiota yhden kielen ja standardisoinnin avulla. Pohjoismaissa soveltavan terminologian tärkein ala eli sanastotyö on tuottanut runsaasti eri alojen terminologisia sanastoja. Edellä esitetty terminologian historian kuvaus on tarkoituksella suurpiirteinen, sillä aiheesta on kattava katsaus esimerkiksi lähteessä Laurén ja Picht (1993).

Perinteisen terminologisen tutkimuksen tarkoitus on ollut kehittää työkaluja käytännön sanastotyön tarpeisiin (Pilke 2000: 47). Perinteisellä tutkimuslinjalla on pyritty koh-

ti ihannetta. On haluttu selvittää esimerkiksi millainen on hyvä termi. Ihanteellisena sanastotyön tuloksena taas on pidetty käsitelähtöistä sanastoa, jossa käsitteet on määriteltä mieluiten sisältömääritelmien avulla. Standardisoinnin ja kielenkäytön ohjailun avulla kohti ihannetta pyrkivää sanastotyötä sanotaan *normatiiviseksi sanastotyöksi*. Normatiivisesta sanastotyöstä ei ole luovuttu eikä siitä ole edes tarpeen luopua. Yhä edelleen on sovittava suositeltavista termeistä ja käsitteiden määritelmistä väärinymmärrysten välttämiseksi. Niin ikään kielenhuolto on välttämätöntä yhtä hyvin erikoisaloilla kuin yleiskielessä. Terminologian näkökulma on kuitenkin laajentunut, sillä sanastotyötä tehdään nykyisin tekniikan alojen lisäksi monilla sellaisilla erikoisaloilla, joiden termit ja käsitteiden määritelmät soveltuvat huonosti standardisointiin tai joilla sanastotyölle asetetaan muita tavoitteita kuin standardisointi, esimerkiksi käytössä olevan käsitteistön ja termistön kuvaaminen ja selkiyttäminen.

### 3.2 Perinteisen terminologian menetelmät

Perinteinen terminologia pyrkii vaikuttamaan erikoisalojen kieleen kielenhuollon ja standardisoinnin kautta (ks. jakso 3.1). Standardisoinnilla pyritään varmistamaan, että käsitesisältö on sama maasta ja kielestä riippumatta. Esimerkiksi tekniikassa on tarvetta sopia käytettävien osien nimityksistä ja ominaisuuksista, sillä käsite-erojen seuraukset voisivat olla vakavat.

Perinteisen terminologian tärkein työkalu on käsiteanalyysi, jonka avulla selvitetään käsitteiden väliset suhteet ja käsitepiirteet. Käsiteanalyysin lähtökohta on käsite, joka nimetään tietyt vaatimukset täyttävällä termillä. Ihannetapauksessa erikoisalan asiantuntijoista koottu työryhmä laatii käsitteiden määritelmät käsitesuhteiden ja käsitepiirteiden avulla. Työryhmä myös muodostaa termit niille käsitteille, joilta termi puuttuu. Terminologin rooli työryhmässä on toimia terminologian asiantuntijana. Erikoisalatieto on siis työryhmän jäsenten henkistä pääomaa, eikä sitä yleensä tarvitse hakea esimerkiksi teksteistä. Käsiteanalyysin käytännön toteutuksesta on kuvaus esimerkiksi Heidi Suonuutin artikkelissa *Käsiteanalyysi työmenetelmänä* (Suonuuti 1999: 29–42).

Käsiteanalyysi on terminologisessa tutkimuksessa edelleen käyttökelpoinen menetelmä. Erityisen hyvin se sopii tutkimusmenetelmäksi, jos tutkimusaineisto koostuu sanakirjoista ja sanastoista, joissa käsitteiden väliset suhteet ja käsitepiirteet ovat eksplisiittisesti näkyvissä käsitteiden määritelmissä. Tällaiset käsiteanalyysin kannalta ”ihanteelliset” aineistot ovat kuitenkin harvinaisia, sillä sanakirjat ja sanastot ovat usein epäloogisia tai niin suppeita, että tutkimusaineiston koostaminen niistä on työlästä. Sanastoaineistoihin perustuvia terminologisia tutkimuksia on kuitenkin tehty runsaasti (esim. Martin 1992; Pilke 2000), ehkä juuri muunlaisen tutkimusaineiston analyysimenetelmien kehittymättömyyden vuoksi. Erityisen hyvin sanastojen ja sanakirjojen sana-artikkelit sopivat terminmuodostuksen tutkimiseen.

Sanastojen käytöllä terminologisen tutkimuksen aineistona on monia etuja. Tutkittavien käsitteiden valinta ei ole ongelma, koska käsitteitä nimeävät termit ovat sanakirjojen ja sanastojen sana-artikkeleina. Lisäksi määritelmistä näkyvät käsitteiden väliset suhteet ja olennaiset käsitepiirteet, jos käsitteiden määritelmät on laadittu sanastotyön periaatteiden mukaan. Ihanteellisten, rakenteeltaan sanastotyön periaatteiden mukaisten sisältömääritelmien analyysi on jopa mahdollista automatisoida, jolloin manuaalisen työn osuutta voidaan vähentää. Käytännössä tällaisten ihannemääritelmien lukumäärä on kuitenkin niin vähäinen, että automatisointi tuskin nopeuttaa niiden analyysia, sillä tutkimus- ja valistustyöstä huolimatta kaikkia erikoisalojen sanakirjoja ja sanastoja ei ole laadittu hyviä sanastonlaadintakäytänteitä noudattaen.

Käsiteanalyysi on käyttökelpoinen, kun tavoitteena on laatia erikoisalan normatiivinen sanasto tai tutkia esimerkiksi terminmuodostusta tai käsitepiirteitä sanakirja- ja sanastoaineistolla. Niillä on kuitenkin ominaisuuksia, joiden vuoksi ne eivät sovi kaikkien terminologisten tutkimusten aineistoksi. Terminologisessa tutkimuksessa käytetäänkin nykyisin tutkimusaineistona yhä enemmän aitoja käyttötekstejä jo yksin siitä syystä, että läheskään kaikilta erikoisaloilta ei ole laadittu sanastoja kaikilla kielillä. Perinteisen käsiteanalyysin soveltaminen tekstiaineistoon on mahdollista, mutta vaatii runsaasti manuaalista työtä ja perustuu pääosin tutkijan intuitioon. Tekstien käyttäminen terminologisen tutkimuksen tai sanastotyön aineistona edellyttääkin uudenlaisten menetelmien soveltamista. Lähes aina tämä tarkoittaa tietoteknisten apuvälineiden käyttöä manuaalisen työn tukena.

Perinteisen terminologian mukaisen tutkimuksen tavoitteena on kielenkäytön harmonisointi ja normittaminen. Koska tavoitteena ei ole kuvata kielenkäyttöä vaan pyrkiä ohjaamaan sitä, tekstit, joissa terminologinen tieto on hajallaan, piilossa tai jopa ”väärin ilmaistuna” sopivat huonosti perinteisen teorian tutkimusaineistoksi. Tutkimusaineistoa voidaankin pitää eräänlaisena vaikkakin keinotekoisena rajana perinteisten ja uusien tutkimusmenetelmien jaottelussa.

### 3.3 Terminologian tavoitteiden laajeneminen

Standardisointiorientoitunut perinteinen terminologia oli vallitseva 1980-luvulle saakka, jolloin terminologinen tutkimus ja käytännön sanastotyö alkoivat laajeta tekniikan ja luonnontieteiden ulkopuolelle. Laajentuminen toi esiin monia käytännön ongelmia, joihin perinteinen terminologia ja perinteiset menetelmät eivät tarjonneet ratkaisua. Tutkijat ja sanastotyön tekijät huomasivat, että tekniikassa ja luonnontieteissä hyvin toimivat käsiteluoittelut eivät välttämättä sopineet muiden erikoisalojen käsitteiden luokitteluun ja teorian mukainen käsitepiirteiden luokittelu oli riittämätön. Perinteistä teoriaa alettiin syyttää liiasta keskittymisestä standardisointiin ja kielenhuoltoon ja sosiaalisen ulottuvuuden unohtamisesta. Standardisoinnin ja kielenhuollon huumassa unohdettiin, että kieli on altis ympäristön vaikutuksille ja sen voidaan jopa ajatella heijastavan ympäristön muutoksia. Sosiaalisen ulottuvuuden sivuuttamisen vuoksi myös kielen luonnollinen vaihtelu ja muutos ovat saaneet vain vähän huomiota.

1980-luvulla alkanut perinteisen terminologian teoriaan kohdistunut kritiikki on ollut voimakkainta sosioterminologian piirissä. Tämä perinteisen terminologian teorian rinnalle syntynyt suuntaus ei pyri syrjäyttämään aiempaa terminologian teoriaa, vaan painottaa monipuolisempaa näkemystä ja pyrkii kehittämään tutkimusta niin, että myös kielen sosiaaliset aspektit otetaan huomioon. Sosioterminologian kannattajia ovat muun muassa Gaudin (1993) ja Temmerman (2000). Vaikka sosioterminologia on sekin ollut kritiikin kohteena (ks. esim. Myking 2001), se on tuonut terminologiseen tutkimukseen aiemmin sivuutetun sosiaalisen ulottuvuuden ja diakronisen tutkimuksen.

Standardisointiorientoitunut perinteinen terminologia ja sosioterminologia saattavat näyttää vastakkaisilta suuntauksilta, koska sosioterminologian edustajien perinteiseen terminologian teoriaan kohdistama kritiikki on paikoin ollut hyvinkin kärkevää. Suuntauksukset eivät kuitenkaan ole vastakkaisia, vaan ne sopivat hieman erilaisiin tarkoituksiin. Perinteisen terminologian teorian tarjoamat työkalut ovat edelleen käyttökelpoisia muun muassa standardisoinnin tarpeisiin siinä missä sosioterminologian menetelmät sopivat paremmin esimerkiksi humanistisen ja sosiaalialan termistön tutkimukseen.

Nykyisin standardisointi edustaa enää vain yhtä vaikkakin tärkeää osaa terminologisten tutkimusten ja sanastotöiden joukossa. Viime vuosikymmeninä terminologinen tutkimus on lähentynyt ammattikielten tutkimuksen näkökulmaa, jonka mukaan tutki-

muksen tehtävänä on selvittää, millaisia termejä tutkittavassa ammattikielessä käytetään, olivatpa ne hyviä tai huonoja. Sanastotyössä painopiste on siirtynyt ehdottomasta normatiivisuuden vaatimuksesta lähemmäs deskriptiivistä lähestymistapaa. Tällaista erikoisalan kielenkäyttöä kuvaavaa sanastotyötä sanotaan *deskriptiiviseksi sanastotyöksi*. Tavoitteiltaan erilaiset soveltavan terminologian alat, perinteinen normatiivinen sanastotyö ja uudempi deskriptiivinen sanastotyö, täydentävät toisiaan, sillä deskriptiivinen sanastotyö voi olla perustana – ja perusteena – myöhemmälle normatiiviselle sanastotyölle.

Olen tietoisesti välttänyt mainitsemasta sanaa *tiede* terminologian yhteydessä, sillä tiedemaailmassa ei edelleenkään ole yksimielisyyttä siitä, onko terminologia tieteenala vai ei. Usein terminologiaa on pidetty kielitieteen osa-alueena (ks. esim. Leitchik & Shelov 2004: 25), mutta muun muassa Felber (1984) ja Grinčev (1993) pitävät terminologiaa tieteenä. Ottamatta kantaa puolesta tai vastaan totean, että rajat tieteenalojen välillä ovat joka tapauksessa alkaneet hämärtyä monitieteisten tutkimusten yleistyttyä. Perinteisesti terminologian teoria on ammentanut filosofiasta, ennen kaikkea logiikasta, jossa käsitteet ovat keskeinen tutkimuskohde. Uusimmat terminologiset tutkimukset ovat lähes aina monitieteisiä, sillä terminologian teorian lisäksi niiden teoreettisena perustana ovat esimerkiksi heuristiikka, informaatiotieteet, kognitiotieteet tai psykolingvistiikka. Terminologisten menetelmien automatisointi yhdistää terminologian entistä läheisemmin informaatiotieteisiin ja tietokone-lingvistiikkaan. Yhteys informaatiotieteisiin on ymmärrettävä, onhan terminologiassa kyse tiedon siirrosta erikoisalaviestinnän kautta. Lähimpänä viestintätieteitä on kommunikatiivinen terminologia (Feliu & Cabré 2002), jossa korostetaan terminologian kommunikatiivista tehtävää. Kategorisaation tutkimus puolestaan on yhteistä terminologiassa ja muun muassa sosiaalipsykologiassa, lingvistiikassa (esim. Lakoff 1987) ja kognitiivisessa psykologiassa (esim. Rosch 1978).

### 3.4 Uudet terminologiset menetelmät

Terminologian tutkimuskohteista termeillä ja käsitetiedolla voidaan ajatella olevan kolme tasoa, jotka ovat kognitiivinen, lingvistinen ja kommunikatiivinen (vrt. Sager 1990: 13). Kognitiivinen taso kuvaa sitä, miten tieto rakentuu ja järjestyy ihmisen mielessä eli ajatusprosesseissa. Koska ihmisen ajatusmaailma on hankala tutkimuskohde ainakin terminologisilla tutkimusmenetelmillä, kognitiivisen tason eli käsitteiden tutkimus jää odottamaan kehittyneempiä menetelmiä. Lingvistinen taso kuvaa sitä, miten kognitio puetaan sanoiksi eli miten käsitteitä määritellään ja nimetään. Lingvistisen tason ilmiöitä säätelevät yleiset kielen periaatteet ja mekanismit (Ivina 2003: 168). Terminologian tehtävä on järjestää kognitiivisten prosessien tuottama tieto käsitejärjestelmiksi ja kielen tasolla termijärjestelmiksi. Termejä ja määritelmiä on tutkittu runsaasti, koska lingvistisen tason tutkimusmenetelmät ovat olemassa. Sen sijaan kommunikatiiviseen tasoon on tutkijoiden kiinnostus herännyt vasta viime aikoina (ks. esim. Feliu & Cabré 2002). Kommunikatiivisella tasolla kuvataan sitä, miten käsitteistä viestitään, miten termejä käytetään ja mitä määritelmätietoa käsitteestä annetaan. Kommunikatiivisen tason tutkimuksessa on otettava huomioon myös viestintätilanne, sillä tutkittava teksti on kirjoitettu jotakin viestintätehtävää varten eikä se voi olla vaikuttamatta tekstissä annettavaan tietoon. Terminologisesta näkökulmasta tiedon kiinnostavuuteen ja luotettavuuteen vaikuttavat esimerkiksi lähettäjä ja vastaanottaja sekä näiden tiedon taso. Kommunikatiivisen tason ilmiöitä säätelevätkin ennen kaikkea erikoisalan tarpeet (Ivina 2003: 168). Vaikka keskityn tässä tutkimuksessa lingvistiseen tasoon, pyrin ottamaan muutkin tasot huomioon, sillä usein lingvistisen tason ilmiöitä selittäviä tekijöitä on syytä etsiä kognitiiviselta tai kommunikatiiviselta tasolta. Kaikki kolme tasoa ovat



keskenään vuorovaikutussuhteessa, sillä kommunikatiivinen taso on lähtökohta, jolta sanastotyötä tekevä terminologi tai kääntäjä lähtee pyrkiessään kohti lingvististä tasoa, jolla kognitio on puettu sanoiksi.

Historiallisista syistä terminologinen tutkimus on kohdistunut lähinnä luonnontieteiden ja tekniikan aloihin ja aineellisiin tarkoituksiin (Pilke 2000: 135). Viime aikoina terminologista tutkimusta ja sanastotyötä on alettu tehdä muilla kuin perinteisesti terminologien kiinnostuksen kohteena olleilla tekniikan ja luonnontieteiden aloilla. Esimerkiksi Suomessa sanastotyötä tekevä Tekniikan Sanastokeskus muutti nimensä Sanastokeskus TSK:ksi, sillä jo ennen nimenmuutosta keskus oli hyvän aikaa tehnyt sanastotyötä esimerkiksi sosiaalialalla ja ympäristöalalla. Näiden terminologisesta näkökulmasta uusien alojen erikoiskieli poikkeaa tekniikan ja luonnontieteiden kielestä, eivätkä perinteiset menetelmät sellaisenaan toimi. Uudet erikoisalajat ovat nostaneet esiin uusia ongelmia sekä aineiston valinnassa että sen analyysissä. Uusien alojen myötä tutkijat ovat heränneet pohtimaan sellaisia tutkimuskysymyksiä, jotka eivät perinteisesti ole olleet relevantteja. Esimerkiksi terminologisen tiedon diakroniseen tutkimukseen sovellettavat menetelmät, kuten käsitteiden muutoksen ja termivariaation tutkimusmenetelmät, voidaan laskea uusiin menetelmiin, koska perinteisesti terminologisessa tutkimuksessa on sovellettu synkronian periaatetta, jonka mukaan vain tutkimushetken tilanne on kiinnostava. Diakronista tutkimusta on toistaiseksi tehty vähän, mutta moni tutkija on todennut, että tällainen tutkimus on tarpeen ja saattaa antaa mielenkiintoista tietoa (esim. Temmerman 2000: 230).

Kuten mainitsin jaksossa 1.4, aitojen tekstien käyttö tutkimusaineistona ei ole harvinaista. Aidot käyttötekstit ovat tutkijan tai sanastotyön tekijän paras tietolähde, jos tavoitteena on esimerkiksi tutkia erikoisalan termien variaatiota. Jos tarkasteltavasta erikoisalasta ei ole laadittu sanakirjoja tai sanastoja, ne ovat vanhentuneita tai eivät noudatta hyviä sanastonlaadintakäytänteitä, ei terminologinen tutkimus tai sanastotyö onnistu ilman aidoista teksteistä koottua aineistoa.

Tekstien käyttö terminologisen tutkimuksen ja sanastotyön aineistona tuo tutkijoiden ja sanastotyön tekijöiden eteen uudenlaisen ongelman, johon ei ole terminologian teoriassa kiinnitetty suurta huomiota kuin vasta viime vuosikymmeninä ja johon ei vielä ole löydetty tyydyttävää ratkaisua. Tämä ongelma on terminpoiminta. Koska terminologisessa tutkimuksessa on usein käytetty aineistona valmiita standardeja ja sanastoja, ei termien poimintamenetelmien kehittäminen ole ollut tutkijoiden ensisijainen tavoite. Terminpoiminnan automatisointi ei ennen tietotekniikan kehittymistä olisi ollut mahdollistakaan, sillä ennen 1980-luvun loppupuolta ei tietokoneiden kyky riittänyt suurten tekstimassojen käsittelyyn (Ahmad & Rogers 2001: 727).

Terminpoimintaan liittyy toinenkin kysymys, joka on sivuutettu perinteisessä terminologiassa. Perinteisen terminologian teorian mukaan termin käyttö ei riipu kontekstista (Sager 1990: 10). Viestintätilanne on kuitenkin ehdottoman tärkeä, ja vain kontekstia tarkastelemalla voidaan sanoa, ”käyttäytyykö kieli terminologisesti vai normaalisti”, kuten Pearson (1998: 26) asian ilmaisee. Terminpoiminnassa onkin mietittävä ennen kaikkea sitä, miten tai millaisessa ympäristössä termiä käytetään. Terminologisen tutkimuksen näkökulmasta sanakirjojen, sanastojen ja standardien puute on, että käsitteet ja termit ovat niissä tavallaan kivettyneinä ja irrallaan käyttöympäristöstään. Päästäkseen standardiin tai sanakirjaan niiden on täytynyt jo vakiintua käyttöön. Ne myös pysyvät standardeissa ja sanakirjoissa useita vuosia, sillä standardien tai painettujen sanakirjojen julkaisu on pitkä prosessi (standardin laadinnan vaiheista ks. esim. [www.sfs.fi](http://www.sfs.fi)). Sanakirjoissa ja sanastoissa saattaa siis olla vanhentuneita termejä tai määritelmiä.

Termien tutkimusta on yleensä harjoitettu irrallaan niiden nimeämien käsitteiden tutkimuksesta, vaikka termi ja käsite liittyvät kiinteästi yhteen. Aitojen tekstien käytön

ja diakronisen tutkimuksen ohella eräs viimeisin terminologisen tutkimuksen suuntaus onkin käsitteen ja termin rajan häivyttäminen. Termiä ei enää nähdä kivettyneenä kielen yksikkönä, vaan elävän kielen osana, jolla on muuntautumiskykyä, haluttiinpa sitä tai ei. Termien ja käsitteiden sijasta jotkut tutkijat puhuvat mieluummin esimerkiksi terminologisista fraseeista (terminological phrasemes; ks. Montero-Martínez ym. 2002) tai ymmärryksyksistä (unit of understanding; Temmerman 2000). Puhuessaan terminologisista fraseeista Montero-Martínez ym. haluavat laajentaa termin käsitettä siten, että syntaktisten kaavojen sijasta tarkastellaan ontologisia kaavoja, joiden mukaan terminologiset lausekkeet muodostuvat. Näkökulman muutos vie terminologista tutkimusta lähemmäs ontologiaa, joka tutkii ympäröivää todellisuutta.

Uusin terminologinen tutkimus on lähes aina monitieteistä, ja terminologit käyttävät nykyisin terminologisten menetelmien lisäksi lähialojen tutkimusmenetelmiä. Usein näin tapahtuu käytännön pakosta, sillä terminologiassa ei aina ole tutkimuskysymyseen sopivia tai toimivia omia menetelmiä. Menetelmiä on lainattu esimerkiksi kognitiotieteistä, jotka tutkivat tiedon käsittelyä ja tallennusta ihmisen muistissa. Terminologian kiinnostuksen kohde, erikoisalan käsitteet ja niitä nimeävät termit, ovat monimutkaisten kognitiivisten prosessien tulosta, joten yhteys terminologian ja kognitiotieteiden välillä on selkeä. Esimerkiksi Venäjällä kognitiivinen terminologia kehittyikin tällä hetkellä nopeasti (ks. esim. Ivina 2003; Manerko 2003).

Uusilla terminologisilla tutkimusmenetelmillä pyritään ratkaisemaan ongelmat, joiden voittamiseen aiemmat menetelmät ovat osoittautuneet riittämättömiksi. Haasteita tutkimusmenetelmien kehittäjille ovat esimerkiksi termien ja käsitetiedon poiminta teksteistä ja tiedon järjestäminen niin, että se on helposti käytettävissä (ks. esim. Pearson 1998: 124). Tähän liittyy myös terminologisen tutkimuksen ja sanastotyön joidenkin työvaiheiden automatisointi. Tutkimukseni pyrkiikin tuomaan teoreettista tietoa automaattisten menetelmien kehittämisen perustaksi.

### 3.5 Erikoiskielet terminologian näkökulmasta

Kuten aiemmin totesin, terminologian tutkimuskohde on erikoisalojen käsitteet ja erikoiskielten termit. Erikoisalan käsitteitä ja erikoiskielten termejä voi tutkia monella tavalla, mutta nykyisin tietyn erikoisalan termejä ja käsitetietoa on parhaiten tutkijoiden saatavilla tarkasteltavan erikoisalan julkaisuissa. Jotta tutkimusaineiston valinta ja analyysi onnistuisivat, on ensin perehdyttävä alan erikoiskieleen. Haarala (1981: 11) toteaaakin, että ammatillista erikoiskieltä voidaan pitää työkaluna, jonka avulla erikoisalan käsitemaailmaa jäsennetään, kuvataan ja muokataan. Ensimmäinen kysymys on, täytääkö tarkasteltavan alan kieli erikoiskielen kriteerit eli voidaanko tämän erikoisalan kielestä tutkia haluttuja erikoiskielen piirteitä. Tutkimuskohteen ymmärtämiseksi on siis tiedettävä, mikä on erikoiskieli, käsite ja termi. Tässä jaksossa keskityn erikoiskielten ominaisuuksiin, sillä käsitettä ja termiä olen pohtinut luvussa 2.

Erikoiskieli määritellään kielimuodoksi, jota käytetään viestinnässä tietyllä erikoisalalla (Terminologian sanasto 2006, s.v. *erikoiskieli*). Tämä määritelmä rajaa erikoiskielen käytön tietylle erikoisalalle. Muita erikoiskielen kriteereitä ovat kielentutkijoiden mukaan alakohtaiset erityisilmaukset (lähinnä termit), erikoistuneet sosiaaliset puitteet, tietty käyttäjäryhmä, tietty kielen omaksumistapa, kielelle ominaiset teemat ja sanonnat sekä semanttiset ja rakenteelliset poikkeamat yleiskielestä (ks. Ahmad 1994: 3; Karhalme 1996: 36; Niemelä 2003: 29; Picht & Draskau 1985: 20–21; Sager 1994: 29). Niemikorpi (1996: 7–12) jaottelee erikoiskielet ammattikieliin, jotka rakenteellisesti pohjaavat kirjakielen normeihin, ammattislangeihin, jotka ovat puhekielistä ammattikieltä, vanhoihin käsityöläiskieliin ja populaaristettuun eli yleistajuistettuun kieleen,

joka on itse asiassa jo osa yleiskieltä. Toinen yleisesti käytetty jaottelu on erikoiskielten jako ammattikieliin ja harrastekieliin (esimerkiksi golf; ks. esim. Vehmas-Lehto 2005: 93). Eri lähteissä tarkoitetaan useimmiten todennäköisesti ammattikieliä, vaikka puhutaan erikoiskielistä. Eräissä kielissä korostetaan erikoiskielten ammatillista tehtävää (Fachsprache; language for special purpose; fackspråk). Tämä terminologinen epätarkkuus pätee myös tähän työhön: kun puhun erikoiskielistä, tarkoitan nimenomaan ammattikieliä, niin kuin Niemikorpi on ne määritellyt.

Käytännössä erikoiskielen ja yleiskielen välinen raja on veteen piirretty viiva, sillä erikoiskieli perustuu yleiskieleen (Haarala 1981: 9). Kielitieteilijöiden välillä onkin erimielisyyttä yleiskielen ja erikoiskielen suhteesta. Ovatko ne osittain samoja, onko toinen toisen osa vai ovatko ne täysin erillisiä? Esimerkiksi Grinëv (1993: 26) on sanonut, että erikoiskielen termi on ennen kaikkea yleensä kieleen kuuluva sana tai sanaliitto ja vasta toissijaisesti erikoiskielen ilmaus. Yleiskieli määritellään kielimuodoksi, joka on kaikille tietyn kielen puhujille yhteinen. Termit kuuluvat kuitenkin vain tietyn ammattialan tai harrastajapiirin erikoiskieleen, eivät kaikille yhteiseen yleiskieleen. Yleiskielessä puolestaan on ilmauksia ja piirteitä, joita erikoiskielissä ei tavata. Tällä perusteella yleiskieli ja erikoiskieli ovat vain osittain samoja ja kummassakin on aineksia, joita toisessa ei ole. Väitetä, että yleiskieli on osa erikoiskieltä, voidaan tosin ainakin teoriassa perustella sillä, että erikoiskielet ovat sanastoltaan laajempia kuin yleiskieli.

Yleiskielen ja erikoiskielen täydellistä erottamista perustellaan niiden erilaisilla käyttötilanteilla (Bergenholtz 1995: 16–18). Esimerkiksi Ahmad ja Salway (1997: 95) erottavat luonnolliset kielet (natural languages) erikoiskielistä (specialist languages; esimerkiksi tieteen kieli) ja sanastoltaan ja lauserakenteeltaan rajoittuneista operaatiokielistä (operational languages; esimerkiksi merenkulussa käytetty *Seaspeak*). Ahmadin ja Salwayn jaottelun mukaan erikoiskielet eivät siis kuulu luonnollisiin kieliin. Tämä näkemys on kuitenkin hieman erikoinen, joten tässä luokittelussa luonnollisella kielellä ilmeisesti tarkoitetaan yleiskieltä. Jos erikoiskielet ja operaatiokielet eivät ole luonnollisia kieliä, ne ovat keinokieliä. Erikoiskieli muuttuu keinokieleksi kuitenkin vasta, kun se on menettänyt kaikki yleiskielen piirteet (Sager 1994: 47). Erityispiirteistään huolimatta sekä erikoiskielet että operaatiokielet suurimmaksi osaksi noudattavat yleiskielten syntaksia ja käyttävät niiden sanastoa kielestä riippumatta. Erikoiskielille tärkeä terminmuodostus noudattaa sekin yleiskielten sananmuodostussääntöjä, ainakin tämä on ihanne. Näin erikoiskielet ja operaatiokielet ovat luonnollisia kieliä erityispiirteistään huolimatta.

Bergenholtz (1995: 18–19) esittää vaihtoehdon, jossa yleiskieli ja erikoiskieli menevät osittain päällekkäin. Yhteinen alue on sanastoa, jota erikoisanalan sanakirjoissa ei yleensä ole. Sanastoltaan, käyttöalaltaan ja vaihtelun siedoltaan rajoittuneet operaatiokielet on tarkoituksenmukaista erottaa omaksi kielimuodokseen erikoiskielten sisällä. Hyvä esimerkki operaatiokielestä on Ahmadin ja Salwayn mainitsema merenkulun turvallisuuden tarpeisiin kehitetty *Seaspeak*, jonka avulla pyritään tehokkaaseen ja virheettömään viestintään.

Niemikorven (1996: 100) mukaan erikoiskielet voidaan yleensä tunnistaa siitä, että niiden ymmärtämiseen vaaditaan erikoisanalan koulutusta tai muuten hankittua erikoisanalan tietämystä. Erikoiskieli on siis eräällä tavalla sisäpiirin kieli, jonka käyttäjäksi tullaan ehkä vasta vuosien opiskelun jälkeen. Ahmad (1994: 3) sisällyttää erikoiskieltä käyttävään yhteisöön erikoisanalan asiantuntijat, alan opiskelijat, hallintohenkilöt, tekniset kirjoittajat, terminologit ja kääntäjät.

Terminologian kannalta tärkein erikoiskielten tunnusmerkki on termit (Ahmad & Salway 1997: 98). Juuri termit tekevät erikoiskielestä vaikeasti ymmärrettävän niille lukijoille tai kuulijoille, jotka eivät ole sisäpiiriläisiä eli eivät ole päässeet erikoiskieltä



käyttävän yhteisön jäseniksi koulutuksen tai muuten hankitun erikoisalatietämyksen kautta. Miten termien runsaus sitten näkyy erikoiskielessä? Niemikorpi (1996: 43–44) on verrannut Oulun yliopistoon 1960-luvulla kootun tekstikorpuksen avulla yleiskieltä ja erikoiskieliä. Hänen tutkimuksensa mukaan erikoiskielen tekstissä on muun muassa enemmän itsenäisen merkityksen sisältäviä nomineja (substantiiveja, adjektiiveja ja numeraaleja), yhdyssanoja ja keskenään rinnasteisia lauseenjäseniä kuin yleiskielen tekstissä. Sen sijaan verbejä, viittaus- ja suhdeseanoja sekä lauseiden välistä rinnastusta on vähemmän. Erikoiskielen tekstissä on myös keskimäärin pidempiä sanoja, lauseita ja virkkeitä kuin yleiskielen tekstissä.

Terminpöiminnän kannalta merkittävin on Niemikorven tutkimustulos, jonka mukaan erikoiskielissä substantiivit ja yhdyssanat ovat yleisempiä ja verbit harvinaisempia kuin yleiskielessä. Tämä näyttää olevan erikoiskielen ominaisuus kielestä riippumatta, sillä erikoiskieliä tutkineet Ahmad ja Rogers (2001: 726) ovat todenneet, että erikoiskielen teksteissä on enemmän yhdyssanoja (compound noun), monikollisia sanoja ja substantivoituneita verbejä (nominalized verbs) kuin yleiskielen teksteissä (ks. myös Ahmad 1994: 3, 5, 6.). Esimerkiksi englannin yleiskielen Longman-korpuksessa 100 yleisimmän sanan joukossa on vain kaksi substantiivia (*man, time*). Vertailun vuoksi mainittakoon, että Ahmadin ja Salwayn turvallisuuskielen tutkimuksessa vertailukorpuksena käytetyissä säteilyfysiikan (laajuus 85 109 sanaa) ja säteilyturvallisuuskorpuksessa (laajuus 189 169 sanaa) 100 yleisimmästä sanasta noin 40 prosenttia oli substantiiveja. (Ahmad & Salway 1997: 98–99.) Jos nämä erikoiskielen ominaisuudet yhdistetään tietoon, että erikoiskielessä on runsaasti termejä, päädytään johtopäätökseen, jonka mukaan termit ovat todennäköisesti nomineja ja yhdyssanoja tai sanaliittoja kielen rakenteesta riippuen. Niemikorven mainitsema keskenään rinnasteisten lauseenjäsenten runsaus saattaa puolestaan viitata termiluetteloihin.

Hämmästyttävää on, että sen paremmin Niemikorpi kuin Ahmad ja Rogers eivät mainitse kansainvälisiä lainoja erikoiskielen tunnusomaiseksi piirteeksi. Englannin erikoiskieliä koskeissa tutkimuksissa tämä on ymmärrettävää, koska juuri englanti on tavallisesti lainoja antava kieli, mutta ainakin suomen ja nykyisin myös venäjän erikoiskielissä lainasanoja todennäköisesti viljellään huomattavasti enemmän kuin yleiskielessä. Syy lainasanojen runsauteen löytyy erikoiskielen perimmäisestä tarkoituksesta eli siitä, että erikoiskielet ovat kehittyneet palvelemaan ennen kaikkea erikoisalan eivätkä yleiskielen tarpeita (Niemi 2003: 37). Kansainvälistä yhdenmukaisuutta pyritään erikoiskielissä jopa tietoisesti lisäämään harmonisoinnin avulla. Kansainvälisyys näkyy erikoiskielissä termistön tasolla, mutta käsitetasolla se on joillakin aloilla vielä silmiinpistävämpää. Siinä missä yleiskielen sanojen merkitys riippuu kontekstista, kansainvälisen erikoisalan käsitteiden käsitesisältö saattaa säilyä muuttumattomana yli kieli- ja kulttuurirajojen. Poikkeuksena ovat kulttuurisidonnaiset erikoiskielet, kuten lain ja hallinnon kieli tai sosiaalialan kieli. Merenkulun kieli ei kansainvälisyytensä vuoksi ole kulttuurisidonnainen, mutta kansallisilla merenkulun kielillä, kuten merenkulun suomen- ja merenkulun venäjällä voi tästä huolimatta olla kulttuurisidonnaisia piirteitä.

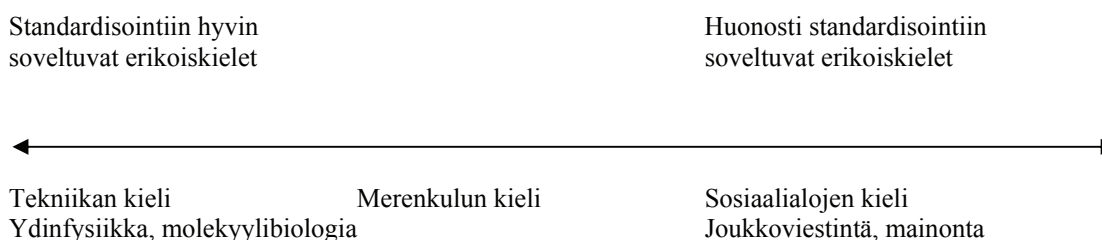
Sanastollisten ja syntaktisten erojen lisäksi erikoiskielet eroavat yleiskielestä pragmaattisella tasolla, koska erikoiskielen tehtävä on palvella nimenomaan tiettyä toimintaa: erikoisalojen viestintää (vrt. Niemi 2003: 31). Näin erikoiskielillä on toinen aihepiiri, viestintätilanne ja käyttäjäryhmä kuin yleiskielellä (Sager 1994: 28). Eri viestintätilanteet vaativat erikoistumistasoltaan erilaisia ilmaisuja, sillä erikoiskielellä viestitään esimerkiksi asiantuntijalta asiantuntijalle tai asiantuntijalta maallikolle. Viestintätilannetta, jossa erikoiskielellä viestitään tietyssä sosiaalisessa ympäristössä tiettyä tarkoitusta varten ja viestintään osallistuu rajoitettu joukko viestijöitä, voidaan sanoa myös erikoiskielen sosiaalisesti ulottuvuudeksi (Picht & Draskau 1985: 11). Viestintätilanteen

lisäksi erikoiskielen muusta kielenoppimisesta poikkeavaa omaksumistapaa voidaan pitää sosiaalisen ulottuvuuden osana.

Edellä olen tarkastellut erikoiskieliä yhtenä luonnollisen kielen muotona, ikään kuin eri tieteenalojen, ammattialojen ja harrastealojen kielet olisivat yksi homogeeninen ryhmä. Näin ei kuitenkaan ole, vaan erikoiskielten välillä on selkeitä eroja. Ensinnäkin erikoiskielten termien määrä voi vaihdella huomattavasti. Ahmad (1994: 3) esittää kielitieteen ja fonetiikan englanninkielisten termien määräksi 2000 ja sähkötekniikan termien määräksi 2 000 000. Luvut perustuvat hakusanojen ja tietueiden määrään eri alojen sanakirjoissa ja termipankeissa.

Toiseksi kunkin erikoisalan kommunikaatiotarpeet ovat johtaneet alakohtaisten kielellisten konventioiden syntymiseen, sillä kuten Niemikorpi (1996: 110) toteaa, erikoiskielet eivät ole kehittyneet ensisijaisesti kansalliskielten muutos- ja kehityspaineen syysäminä vaan kunkin erikoisalan kehittymisen seurauksena. Suomessa erikoiskielissä tosin on myös kansalliskielisyyteen ohjaava komponentti, sillä anglismit istuvat huonosti suomeen. Sijapäätteiden liittäminen englanninkielisiin sanoihin ei ole luontevaa, eivätkä vierasperäiset konsonantit sovi suomen kielen äänneympäristöön. Erikoisalojen alakohtaiset kielelliset konventiot vaikeuttavat terminologisen tiedon poiminnan automatisointia. Esimerkiksi kielelliset ilmaukset, joiden avulla teksteistä voi poimia termejä ja käsitetietoa, ovat alasta riippuvia. Tähän ongelmaan palaan luvussa 7.

Erikoiskiellen voidaan Riggsin (1993: 216) tavoin ajatella olevan jatkumo, jonka toisessa päässä ovat standardisointiin hyvin soveltuvat erikoiskielet ja toisessa päässä huonosti standardisointiin soveltuvat erikoiskielet. Ääripäiden väliin jäävät ne erikoiskielet, joissa standardisointi on mahdollista, vaikka saattaa olla vaikeaa (ks. kuvio 4).



KUVIO 4. Erikoiskiellen soveltuvuus standardisointiin

Tähän saakka terminologisia menetelmiä on kehitetty lähinnä standardisointiin soveltuvien erikoiskiellen tutkimukseen. Standardisointiin soveltuvien erikoiskiellen joukkoon on tapana lukea tekniikan erikoisala. Terminologin kannalta hankalimpia tutkimuskohteita ovat huonosti standardisointiin soveltuvat erikoiskielet. Hankalasti standardisoitavissa erikoiskielissä käytetään esimerkiksi paljon metaforia, joihin on terminologiassa alettu kiinnittää huomiota vasta 2000-luvun taitteessa (ks. esim. Temmerman 2000). Esimerkiksi sosiaalialojen kieli soveltuu huonosti standardisointiin, sillä Riggsin (1993: 216) mukaan terminologinen standardisointi on sosiaalitieteilijöille lähes kirosana. On myös mahdollista, että standardisointi kohdistuu vain tiettyyn osaan erikoiskieltä. Esimerkiksi ilmailuliikenteessä ja meriliikenteessä on erikoisalan sisäisten standardisointipaineiden vaikutuksesta kommunikoinnissa otettu käyttöön tietyt fraasit<sup>1</sup>, joista on

<sup>1</sup> En käsittele meriliikenteen kansainvälisiä fraaseja tarkemmin, vaikka niiden tavoitteena on nimenomaan merenkulun turvallisuuden parantaminen, vaan keskityn pelkästään termeihin.

sovittu kansainvälisesti. Aloite standardisoimistyölle tulee aina erikoisalan sisältä. (Ks. Riggs 1993: 216.)

Kansainvälisyys on ollut merenkululle ominaista niin kauan kuin sitä on harjoitettu. Merenkulun erikoiskielet ovatkin saaneet vaikutteita eri aikoina merten herruutta pitäneiden maiden kielestä. Nykyisin merenkulun *lingua franca* on kiistämättä englanti, mutta merenkulun englannin termistö perustuu lainoihin hollannin kielestä, sillä hollantilaiset alukset hallitsivat maailman meriä noin sata vuotta 1600-luvun alusta 1700-luvun alkuun, jolloin herruus siirtyi Isolle-Britannialle. Suuri osa hollantilaisista lainoista on päätynyt merenkulun kansainväliseen termistöön. (Bondarenko 1992: 5, 15–16.) Merenkulun suomi on osaksi vanhaa 1800-luvulla käytettyä merimieskieltä. Lainasanoja on paljon, vaikka niitä ei aina mielletä lainoiksi. (Niemi 1996: 150.) Sanaston lisäksi englannin kieli vaikuttaa merenkulun suomeen – ja nykyisin myös merenkulun venäjään – oikeinkirjoituskäytännön kautta. Tutkimusaineistossani onkin tekstejä, joissa englannin kielen vaikutus näkyy niin sanastossa kuin oikeinkirjoituksessakin. Olen ottanut nämä tekstit aineistoon, koska tavoitteeni on tutkia merenkulun suomea ja merenkulun venäjää todellisissa kielenkäyttötilanteissa.

#### 4 TERMEN JA KÄSITETIEDON POIMINTAMENETELMÄT

Suunnitteluvaiheen jälkeen ensimmäinen varsinainen työvaihe sanastotyössä on termininventaarior eli terminpoiminta käytössä olevista lähteistä. Tässä työvaiheessa sanastotyötä tekevä terminologi tai erikoisalan tekstiä analysoiva kääntäjä, tulkki, tekninen kirjoittaja tai opettaja saattaa jossakin tapauksessa tyytyä pelkkään terminlistaan, mutta yleensä tavoitteena on koota tekstistä mahdollisimman paljon myös erikoisalan käsitteitä koskevaa tietoa. Aitojen käyttötekstien kanssa työskentelevän terminologin tai kääntäjän on ensin paikallistettava tekstistä käsitteet, joita koskeva tieto kiinnostaa. Ennen terminpoimintaa on ratkaistava, poimitaanko käsitteiden nimityksiä vai termejä, sillä terminlista ei ole sama kuin käsitteitä nimeävien lausumien lista. Käsitteeseen saatetaan viitata elliptisellä ilmauksella, joka itse asiassa onkin yläkäsitteen nimitys. Esimerkiksi käsitteeseen *meriliikenteen ohjausjärjestelmä* saatetaan viitata yläkäsitteen nimityksellä *liikenteen ohjausjärjestelmä*. Lisäksi monisanaisen termin osa saattaa olla jonkin toisen käsitteen nimitys. Esimerkiksi sanaliittotermissä *meriliikenteen ohjausjärjestelmä* on myös termi *meriliikenne*. Toisaalta käsitteisiin voidaan viitata muuten kuin termeillä tai muilla käsitteen nimityksillä: pronomineilla, määritelmillä tai selitteillä. Aina ei ole helppoa ratkaista, mikä on termi ja mikä ei. Esimerkiksi rajanveto termin ja ammattislangiin kuuluvan nimityksen välillä on usein ongelmallista. Käsitteiden paikallistaminen on silti vaikeampaa kuin termien poiminta, koska tällöin on ratkaistava synonyimiaan ja variaatioon liittyvät ongelmat. Näitä ongelmia käsittelen jaksossa 5.4. Käsitteitä on kuitenkin vaikea paikallistaa ilman terminpoimintaa. Tästä syystä tutkijoiden kiinnostus on kohdistunut terminpoiminnan ongelmiin, sillä niihinkään ei ole vielä löytynyt tyydyttäviä ratkaisuja.

Termien ja käsitetiedon poiminnan avuksi on kehitetty erilaisia menetelmiä. Käytettävän menetelmän valinta riippuu ensinnäkin käytettävästä aineistosta. Terminologisissa tutkimuksissa on perinteisesti käytetty sanakirja- ja sanastoaineistoja, mutta muun muassa terminologit ja kääntäjät ovat jo kauan käyttäneet tekstejä terminologisen tiedon lähteinä. Toiseksi menetelmän valintaan vaikuttaa poiminnan tavoite. Jos keskitytään pelkästään termeihin, on ehkä tarkoituksenmukaista poimia aineistosta kaikki termit. Jos taas tavoitteena on tarkastella keskeisiä käsitteitä, poimitaan vain niitä nimeävät termit. Vaikka sanastotyön lähtökohtana pidetään yleisesti käsitettä (ks. Sanastotyön käsikirja 1988: 145), käytännössä sanastotyökin alkaa termiehdokkaiden poiminnasta (Soininen 1998: 19). Termiehdokkaiden valinnan jälkeen on kuitenkin palattava käsitteisiin, koska käsitesisältö ratkaisee, onko valittu termiehdokas termi (Šelov 2001: 2).

Aiemmissa tutkimuksissa on yleensä keskitytty joko termien tai käsitetiedon poimintaan, koska näin tutkimusongelman rajausta helpottuu (ks. esim. Christensen 2000; Kavanagh 1995; Lahtinen 2000; Sewangi 2001; Sisseck 2005; Soininen 1998; Urizar, Ezeiza & Alegria 2000). Termejä ja käsitetietoa on kuitenkin tarkoituksenmukaista käsitellä toisiinsa liittyvänä kokonaisuutena, koska joka tapauksessa tiettyä käsitettä koskeva käsitetieto on pystyttävä yhdistämään käsitettä nimeävään termiin. Käsitetieto on tärkeää, jotta voisi varmistua siitä, mitä käsitettä termi nimeää. Käsitetieto on tarpeen myös käsitteistä selvittelemiseksi, sillä pelkän terminlistan perusteella ei käsitteiden välisistä suhteista tiedetä mitään (Picht & Draskau 1985: 104). Mielestäni käsitteiden nimitysten ja käsitetiedon erottaminen toisistaan johtaa tiedon pirstoutumiseen, ja vain johonkin käsitetiedon osaan keskittymällä kuva erikoisalan käsitteistöä jää vajavaiseksi.

Termien ja käsitetiedon poimintamenetelmät voidaan jakaa kahteen ryhmään sen mukaan, käytetäänkö apuna tietotekniikkaa vai ei. Poimintamenetelmästä, jossa ei käytetä tietotekniikkaa, käytän nimitystä *manuaalinen terminpoiminta*. Menetelmää, jossa

hyödynnetään tietotekniikkaa, nimitän *puoliautomaattiseksi terminpoiminnaksi* enkä automaattiseksi terminpoiminnaksi, sillä toistaiseksi ei ole pystytty kehittämään tietokoneohjelmaa, joka tuottaisi valmiin käsitejärjestelmän tai kelvollisen terminlistan ilman ihmisen tekemää siivousta. Toinen mahdollisuus olisi käyttää tietokoneavusteisesta terminpoiminnasta nimitystä automaattinen termiehdokkaiden poiminta, joka olisi kaikkein lähimpänä tämänhetkistä terminpoiminnan automatisoinnin todellisuutta (vrt. Ahmad & Rogers 2001: 755). Puoliautomaattinen terminpoiminta (semiautomatic term extraction) on kuitenkin vakiintunut käyttöön englanninkielisissä lähteissä. Seuraavaksi tarkastelen nykyisin käytössä olevia termien ja käsitetiedon poimintamenetelmiä lähemmin omista jaksoistaan.

#### **4.1 Terminpoimintamenetelmät**

Terminologisissa tutkimuksissa, joissa aineistona käytetään valmiita sanakirja- ja sanastoaineistoja (esim. Pilke 2000), ei terminpoiminta tuota ongelmia. Sen sijaan sellaisissa tutkimuksissa tai sanastotöissä, joissa aineistona ovat tekstit, joudutaan ratkaisemaan, miten erottaa termit muusta kielellisestä aineksesta. Terminhaku teksteistä ei ole ollut terminologien erityisen kiinnostuksen kohteena ja terminhaun automatisointi on kiinnostanut lähinnä tietokonelingvistejä ja informaatiohaun tutkijoita. Terminologian teoriassakin terminhaun metodologia on hajanainen ja yhtenäiset kriteerit haun tulosten arvioimiseksi puuttuvat (Soininen 1999a: 12–13). Terminhaku ei ole ollut tutkijoiden erityisen mielenkiinnon kohteena mahdollisesti siksi, että perinteinen terminologian teoria on käsitelähtöinen.

Sanastotyössä kohdealan teksteistä poimitaan yleensä ensin kaikki mahdolliset termiehdokkaat, jolloin termiehdokaslista on aluksi pituudeltaan tyypillisesti 2–4-kertainen verrattuna sanastoon valittavien termien listaan. Toinen vaihtoehto on lähteä liikkeelle pienestä käsitejoukosta ja laajentaa sitä sanastotyön edetessä. (Nykänen 1999: 65.) Jälkimmäistä vaihtoehtoa suosittelee muun muassa Gerd (1986: 27). Tässä vaihtoehdossa tarvittavat lähtökäsitteet voivat löytyä muualta kuin teksteistä, esimerkiksi sanastotyön tekijältä itseltään tai erikoisanalan asiantuntijoilta suullisesti. Muitakin mahdollisuuksia on, kuten Pearsonin (1998) tutkimus osoittaa. Hän käytti menetelmää, jossa analysoitiin ensin manuaalisesti kolme eri alan aineistoa ja analyysin perusteella laadittiin termien osien sanaluokkarakennetta kuvaavat termikaavat. Tämän jälkeen koko aineistosta poimittiin termit puoliautomaattisesti termikaavojen avulla. (Pearson 1998: 121.) Pearsonin käyttämässä menetelmässä siis yhdistyvät puoliautomaattinen ja manuaalinen menetelmä, jota seuraavaksi tarkastelen lähemmin.

##### **4.1.1 Manuaalinen terminpoiminta**

Manuaalinen terminpoiminta perustuu ennen kaikkea poimijan omaan intuitioon siitä, mikä on termi. Olipa poimija tutkija, erikoisanalan asiantuntija tai opiskelija, terminpoimintaa säätelevät samat inhimilliset lainalaisuudet. Kun luemme tekstiä, muodostamme siitä oman mentaalisen representaation, johon vaikuttavat muun muassa tieto maailmasta, konteksti ja tavoitteet (Pälli 2003: 51). Mentaalisen representaation muodostumiseen vaikuttavat ennen kaikkea tekstinulkoiset, eivät tekstinsisäiset tekijät, kuten kieliopilliset ominaisuudet tai esimerkiksi sanojen frekvenssit.

Terminologit, kääntäjät ja tulkit ovat terminpoiminnassa käyttäneet yleensä manuaalista menetelmää. Myös tutkijoilla terminpoiminta teksteistä on saattanut perustua tutkijan omaan intuitioon, kuten esimerkiksi muotoilun teorian sanaston termistymistä tutkineella Karihalmeella (1996: 25, 77). Toinen vaihtoehto on pyytää erikoisanalan asiantunti-

joita poimimaan termit. Näin on tehnyt esimerkiksi Fulford (2001). Kolmas, lähinnä vain tutkimustarkoituksiin sopiva vaihtoehto on menetelmä, jossa terminpoiminnassa käytetään koehenkilöitä. Esimerkki tästä menetelmästä on Opitzin (1979: 90–91) vuonna 1978 tekemä koe, jossa merenkulun kansi- ja konepuolen opiskelijat tunnistivat noin 1 000 sanan englanninkielisestä aikakauslehtiartikkelista sanat, jotka heidän mielestään olivat merenkulun erikoisan sanoja (nautical affairs).

Manuaalisella menetelmällä on tiettyjä etuja, jotka puolustavat sen käyttöä edelleen, vaikka puoliautomaattisiakin terminpoimintamenetelmiä on olemassa. Ensinnäkin manuaalisen menetelmän välineiksi riittävät kynä ja analysoitava teksti paperille tulostettuna. Toiseksi tekstiä ei tarvitse erikseen valmistella terminpoimintaa varten. Kolmanneksi manuaalinen terminpoiminta on varsin luotettava menetelmä, jolla tekstistä saadaan poimittua joko kaikki termit tai keskeiset termit tavoitteesta riippuen. Neljänneksi termilistassa ei ole ”ylimääräistä” ainesta, vaan kaikki termilistasta päättyneet termiehdokkaat ovat ainakin poimijan arvion mukaan termejä.

Erityisen hyvin manuaalinen menetelmä sopii terminpoimintaan lyhyistä teksteistä, ja niillä manuaalinen poiminta antaa paremman tuloksen kuin puoliautomaattinen poiminta. Lyhyiden tekstien käsittely käsin on tarkoituksenmukaista myös silloin, kun tekstiä ei ole tallennettu sähköisesti, sillä tallennus vaatii oman aikansa. Pitkien tekstien manuaalinen käsittely voi olla paikallaan silloin, jos terminpoimijan tarkoituksena on tehdä sanasto hänelle ennestään oudosta aiheesta. Tekstiä lukiessaan poimija samalla ammentaa tietoa erikoisalasta ja sen käsitteistä.

Manuaalisen terminpoimintamenetelmän heikkoutena taas on sen hitaus ja ihmis työvoiman tarve. Sanastotyötä tekevä terminologi hyötyy tekstien manuaalisesta tarkastelusta alussa, jos erikoisala on outo, mutta koska sanastotyö vaatii suuren tekstimassan läpikäymistä, tietyssä vaiheessa tekstit alkavat toistaa itseään ja terminologi haluaisi mielellään automatisoida terminpoimintaa.

Hitauden ja sitovuuden lisäksi manuaalisen terminpoiminnan puute on sen subjektiivisuus. Aiemmissa tutkimuksissa on todettu yksimielisesti, että termien tunnistaminen on aina subjektiivista ja tulokset vaihtelevat huomattavasti kokeneidenkin terminpoimijoiden välillä (Laakso 2003: 27; Soininen 1998: 6–7, 1999b: 7). Poimijoiden tulokset saattavat poiketa toisistaan huomattavasti, vaikka poimijoiden ikä, sukupuoli, koulutus ja erikoisan tuntemus olisivat likimain samat. Poimintatulos saattaa vaihdella jopa saman poimijan tekemänä, jos ajankohta ja poimintatilanne muuttuvat. Näihin manuaalisen terminpoiminnan ongelmiin haetaan ratkaisua puoliautomaattisista menetelmistä.

#### **4.1.2 Puoliautomaattinen terminpoiminta**

Tiedon hallinnan automatisointi ja tekstikorpusten käyttö tiedonhaussa on yleistynyt 1980-luvulta lähtien, jolloin sähköisesti tallennettujen tekstikorpusten analyysi korpus työkalujen avulla alkoi olla teknisesti mahdollista. Puoliautomaattisten menetelmien kehitys on ollut nopeaa varsinkin tiedonhaun (information retrieval, IR) alalla, jolla kehitetään menetelmiä indeksoinnin (automatic indexing) ja avainsanojen uuttamisen automatisoimiseksi (automatic keyword extraction). Nämä menetelmät ovat lähellä terminpoiminnan automatisoinnin menetelmiä, mutta termien sijasta niillä pyritään saamaan osumalistaan avainsanoja eli sanoja, jotka mahdollisimman hyvin kuvaavat tekstin sisältöä. Myös korpuslähtöiseen tietokonelingvistiikkaan kuuluva automaattinen termintunnistus (automatic term recognition, ATR) on hyvin lähellä terminologista terminpoimintaa. Kummankin toiminnan tavoite on sama, teknisten termien poiminta erikoisan korpuksesta, mutta tietokonelingvistiikan alalla terminpoimintaa tarkastellaan tekniikan näkökulmasta. Tavoitteena on kehittää niin hyviä tietokoneohjelmia, että poi-



minta on mahdollista kokonaan automatisoida. Terminologisessa terminpoiminnassa taas poimintaa tarkastellaan kielenkäyttäjän näkökulmasta, josta nähtynä tekniikka on vain hyvä apuväline. Vaikka indeksoinnin ja tietopalvelun tarpeisiin kehitetyt hakumenetelmät eivät sellaisenaan olekaan terminologisen tutkimuksen tai sanastotyön tarpeisiin kehitettyjä menetelmiä, niistä voisi olla hyötyä myös sanastotyössä. Niiden soveltaminen terminologiseen tutkimukseen jää kuitenkin tämän tutkimuksen ulkopuolelle, sillä tässä työssä keskityn vain terminpoimintaan tarkoitettuihin tietokoneohjelmiin.

Myös terminpoiminnan avuksi on eri puolilla maailmaa 1980-luvun lopulta alkaen kehitetty tietokoneohjelmia, eikä niistä tänä päivänä ole pulaa (ks. esim. Ahmad & Rogers 2001; Cabré Castellví ym. 2001; Hull 1997: 225–227; Kageura & Umino 1996). Kehitystyöstä huolimatta markkinoilla olevista terminpoimintaohjelmista saadut käyttökokemukset eivät ole olleet pelkästään myönteisiä ja jotkut ohjelmia kokeilleet ovat luopuneet niiden käytöstä ja palanneet manuaaliseen terminpoimintaan. Soinisen (1999a: 59) mukaan syynä ei ole tietoteknisen tai terminologisen osaamisen puute, vaan se, ettei eri alojen osaamista ole osattu yhdistää. Suuret ohjelmistotalot eivät ymmärrettävästi ole halunneet panostaa termien ja käsitetiedon poimintaan tarkoitettujen ohjelmien kehitystyöhön, koska tällaisten ohjelmien markkinat ovat maailmanlaajuisestikin varsin pienet. Terminpoimintaohjelmien kehitystyö jatkunee kieliteknologiayrityksissä ja tutkimuslaitoksissa, sillä konekääntämisessä, sanastotyössä ja tiedonhallinnassa tarvitaan terminpoimintaa (Kageura & Umino 1996: 270; L’Homme ym. 1996: 294).

Puoliautomaattisessa terminpoiminnassa on toistaiseksi käytetty kahta lähtökohdiltaan erilaista menetelmää. Tilastomenetelmä (statistical method) perustuu sanojen frekvensseihin korpuksessa ja lingvistinen menetelmä (linguistic method) sanojen kielellisiin ominaisuuksiin. Myös näiden yhdistelmiä (hybrid methods) käytetään. Selostan seuraavaksi näiden erilaisten puoliautomaattisten termintunnistusmenetelmien periaatteita siinä laajuudessa kuin on tarpeen terminpoimintakokeessa (ks. luku 6) käytettyjen terminpoimintaohjelmien toiminnan ymmärtämiseksi.

#### **4.1.2.1 Tilastolliset puoliautomaattiset terminpoimintamenetelmät**

Tilastollisissa puoliautomaattisissa terminpoimintamenetelmissä termintunnistus perustuu toistoon eli samanlaisena toistuvien merkkijonojen tunnistamiseen ja niiden lukumäärien laskemiseen. Lähtökohdiana on oletus, että terminologiset lausekkeet (terminological phrases) toistuvat tietyn erikoisalan teksteissä useammin samanlaisina kuin muut lausekkeet, joissa on vaihtelua ja erilaisia määriteosia. Termeissä ei määriteosaa yleensä voi vaihtaa muuttamatta käsitettä, jota termi nimeää. (Urizar ym. 2000: 374.) Käänteisesti saman oletuksen voi ilmaista niin, että monisanainen substantiivilauseke, joka ei ole termi, toistuu vain harvoin samanlaisena tekstissä (Soininen 1998: 5).

Tilastollisissa terminpoimintamenetelmissä tutkimusaineistosta tuotetaan ensin listatermiehdokkaista. Ohjelma pisteyttää listaan poimitut termiehdokkaat sen mukaan, kuinka todennäköisesti ne ovat termejä. Pisteiden laskeminen perustuu joissakin ohjelmissa yksinkertaisiin frekvensseihin ja joissakin monimutkaisempiin tilastotieteellisiin laskentakaavoihin, joissa on frekvenssin lisäksi otettu huomioon muitakin tekijöitä.

Tilastollisissa puoliautomaattisissa terminpoimintamenetelmissä poimitaan yleensä ensin yksisanaiset termit ja toisessa vaiheessa yksisanaisten avulla monisanaiset termit (Ahmad & Rogers 2001: 740–743; Sewangi 2001: 10). Sanojen yhdistelmät, jotka esiintyvät tutkittavassa korpuksessa tiettyä raja-arvoa useammin, ovat mahdollisia termejä. Yleisesti lähtökohdiana pidetään olettamusta, että termi esiintyy tekstissä useammin kuin kerran. Lausekkeen frekvenssin on toisin sanoen todettu korreloivan positiivisesti termiyden kanssa (ks. esim. Arppe 1995: 2; Soininen 1998: 7).

Menetelmän etuja muihin puoliautomaattisiin terminpoimintamenetelmiin verrattuna ovat yksinkertaisuus, riippumattomuus kielestä ja alasta sekä kyky poimia yksisanaisia termejä. Menetelmän haitta on epätarkkuus, sillä se tuottaa huomattavan määrän termeiksi kelpaamattomia termiehdokkaita, koska pelkästään sanojen frekvenssiin perustuva menetelmä antaa hakutulokseen yleisiä yleiskielen sanoja (vrt. Ahmad & Rogers 2001: 746). Termeiksi kelpaamattomien termiehdokkaiden määrää voidaan tosin vähentää karsimalla kaikkein taajimmin esiintyvät yleiskielen sanat tuloksesta sulkulistan avulla. Epätarkkuutta esiintyy myös toiseen suuntaan, sillä harvinaiset tai erilaisina muunnelmina esiintyvät termit jäävät termilistan ulkopuolelle (Sewangi 2001: 12). Lisäksi joidenkin tutkimusten mukaan tilastolliset menetelmät antavat tyydyttäviä tuloksia vain lyhyillä teksteillä (L’Homme ym. 1996: 299; ks. myös Soininen 1998: 7).

#### 4.1.2.2 Lingvistiset puoliautomaattiset terminpoimintamenetelmät

Perinteisessä terminologisessa tutkimuksessa on jo kauan tutkittu termejä termikaavojen ja morfologisten ominaisuuksien avulla (Ahmad & Rogers 2001: 746). Tutkimustietoa on viime vuosikymmeninä alettu käyttää puoliautomaattisessa terminpoiminnassa. Lingvistiset terminpoimintamenetelmät perustuvat tiettyjä kieliopillisia rakennekaavoja noudattavien lausekkeiden tunnistamiseen. Ohjelmien termintunnistusalgoritmin taustalla on yleensä oletus, että tunnistettavat lausekkeet ovat 1–4 sanan substantiivilausekkeita ja että jotkin substantiivilausekkeista ovat termejä (ks. esim. Lahtinen 2000: 124). Osa terminpoimintaohjelmista poimii kuitenkin myös sellaisia termiehdokkaita, joissa on adverbi tai prepositio, sekä sanojen yhdistelmiä, joissa on konjunktio (L’Homme ym. 1996: 300). Jotkut ohjelmat sisältävät morfologisen analysaattorin ja pystyvät palauttamaan taivutetut sanat perusmuotoon, jolloin tulostettava termiehdokkaiden luettelo lyhenee.

Didier Bourigault kehitti 1990-luvun alussa terminpoimintamenetelmän, jonka termintunnistusperiaate on käänteinen niille lingvistisille terminpoimintamenetelmille, joissa termintunnistus perustuu tiettyjä kaavoja noudattavien lausekkeiden tunnistamiseen. Bourigaultin LEXTER-ohjelmassa termiehdokkaiksi kelvollisten mahdollisimman pitkien substantiivilausekkeiden (Maximal-Length Noun Phrases) rajat määritellään sulkemalla pois kaikki sellainen kielellinen aines, joka ei voi olla termin osa. Termiä rajaavia kielellisiä aineksia ovat esimerkiksi verbit tai pronominit. Tämän jälkeen ohjelma jakaa pitkät substantiivilausekkeet pääosaan (a constituent in head-position; usein yläkäsitteen nimitys) ja määriteosaan (a constituent in expansion-position). Tämän jälkeen pitkät ja pätkityt termiehdokkaat järjestetään ”termiverkoksi” (”terminological network”) pääsanojen ja määriteosien mukaan. (Bourigault, Gonzales-Mullier & Gros 1996: 771–779.)

Lingvistiseksi terminpoimintamenetelmäksi voidaan sanoa myös suomalaisten tutkijoiden kehittämää NPtool-työkalua, jonka avulla englanninkielisistä teksteistä tunnistetaan substantiivilausekkeet. Terminhaun aloittamista substantiivilausekkeista perustellaan sillä, että termeistä 80–99 prosenttia on substantiivilausekkeita. NPtool tuottaa listan substantiivilausekkeista, jotka ovat mahdollisimman pitkiä. Nämä lausekkeet pilkotaan osiin, jotka kaikki ovat mahdollisia termejä. Jokaisessa lyhyemmässä substantiivilausekkeessa on pääsanana sama substantiivi, mutta määriteosat vaihtelevat. Menetelmää voidaan pitää myös lingvistisen ja tilastollisen menetelmän yhdistelmänä, koska lausekkeiden poiminnan jälkeen lasketaan jokaisen substantiivilausekkeen frekvenssi aineistossa. Substantiivilausekkeet voidaan myös järjestää lingvistisen pääsanan frekvenssin mukaan. Yhdistelmien frekvenssien perusteella terminologi päättelee, mitkä substantiivilausekkeista todella ovat termejä. (Arppe 1995: 1–3.)



Lingvististen menetelmien heikkous on, että teksti on varustettava sanaluokkaa ilmaisevilla leimoilla (tags) tai terminpoimintaohjelman käytössä pitää olla tarkasteltavan kielen kaikki leksikaaliset yksiköt sisältävä ”sanakirja”, johon ohjelma vertaa tekstin sanoja (L’Homme ym. 1996: 297). Leimoista huolimatta lingvistisen menetelmän soveltaminen yksinään tuottaa runsaasti roskaa, koska hakutuloksessa ovat lähes kaikki teksteissä esiintyvät substantiivit, vaikka niistä huomattava osa kuuluu yleiskieleen (ks. esim. Sewangi 2001: 88). Yhden kielen kanssa hyvin toimivaa lingvististä menetelmää ei myöskään voi suoraan soveltaa toiseen kieleen. Lisäksi ohjelmat tunnistavat huonosti yksisanaisia termejä, koska ne hakevat tekstistä vain (substantiivi)lausekkeita, jotka noudattavat tiettyjä syntaktisia kaavoja (L’Homme ym. 1996: 297; ks. myös Dias, Guilloire & Pereira Lopes 2000: 339). Tilastollisten ja lingvististen terminpoimintamenetelmien puutteita on yritetty vähentää kehittämällä niin sanottuja yhdistelmämenetelmiä, joissa nämä kaksi eri tavoin toimivaa menetelmää yhdistyvät.

#### 4.1.2.3 Puoliautomaattiset yhdistelmämenetelmät terminpoiminnassa

Puhtaasti tilastollisten ja lingvististen termintunnistusmenetelmien ohella on myös näiden yhdistelmiä, joissa pääpaino on tavallisesti jommallakummalla menetelmällä (Kageura & Umino 1996: 271). Yhdistelmämenetelmien etu on Lahtisen (2000: 71) mukaan se, että niissä yhdistyy kaksi toisistaan riippumatonta menetelmää, joiden periaate on täysin erilainen. Yhdistelmämenetelmissä termiehdokkaat poimitaan yleensä vaiheittain. Terminpoiminta voidaan aloittaa frekvensseihin perustuvalla tilastollisella menetelmällä, jonka tuottamasta termiehdokaslistasta valitaan tiettyjen rakennekaavojen mukaiset termiehdokkaat. Esimerkiksi Ahmadin ja Rogersin (2001) termintunnistusmenetelmässä termiehdokkaat tunnistetaan ensimmäisessä vaiheessa tilastollisella menetelmällä. Toisessa vaiheessa moniosaiset termiehdokkaat tunnistetaan korpustyökalun konkordanssitoiminnon avulla. Menetelmissä, joissa poiminta aloitetaan yksisanaisista emotermeistä (mother terms), oletetaan, että monisanaiset termit muodostuvat yksisanaisista termeistä tai yksisanainen termi on vähintään monisanaisen termin osa (term head). (Ks. esim. Ahmad & Rogers 2001: 740–749; Kageura & Umino 1996: 273–274; Pazienza 1998/1999: 189)

Mahdollista on myös aloittaa tunnistaminen pitkistä termiehdokkaista, joilla on ennalta määritetty morfosyntaktinen rakenne, ja edetä niistä lyhyempiin. Esimerkiksi C-painomenetelmässä (C-value method) korpusaineistosta tuotetaan ensin monisanaisien termiehdokkaiden luettelo (ks. esim. Frantzi, Ananiadou & Tsujii 1998). Myös sisäkkäisten termien puoliautomaattisessa haussa on pitempi termimuoto haettava ensin (Pearson 1998: 125).

Lingvististä tietoa käytetään suodattimissa, jotka suodattavat sanoja sanaluokan ja sulkulistan (stop-word list) perusteella. Sulkulistalle kootaan sellaiset usein esiintyvät sanat, jotka eivät voi olla termejä tai termin osia (Lahtinen 2000: 117). Tyypillisiä sulkulistan sanoja ovat artikkelit, pronominit ja tietyt adverbis. Sulkulistalle voidaan laittaa myös tekstin metasanat, kuten *kuva* ja *kaavio* (Soininen 1998: 10). Ensimmäisen vaiheen tuottamasta termiehdokaslistasta valitaan lopulliset termiehdokkaat frekvenssin perusteella (Sewangi 2001: 13). C-painomenetelmässä termiehdokkaat järjestetään termiyden mukaiseen järjestykseen, joka perustuu ennen kaikkea termiehdokkaista saatuihin tilastolliseen tietoon. (Mima & Ananiadou 2000: 175–194; Urizar ym. 2000: 375.) Täsäkin menetelmässä lopullisen terminvalinnan tekee ihminen.

Puoliautomaattisia terminpoiminnan yhdistelmämenetelmiä on kehitetty edelleen niin, että termiehdokkaiden morfosyntaktisen rakenteen ja frekvenssin sijasta tai lisäksi menetelmässä saatetaan käyttää erikoisalatunnisteita (domain tags) tai termikontekstisa-

noja (term context words) eli substantiiveja, adjektiiveja tai verbejä, jotka sijaitsevat lauseessa termin läheisyydessä, ennen termiä tai sen jälkeen. Erikoisalatunnisteita ja termikaavoja (term patterns) käytetään esimerkiksi Sewangin (2001) kehittämässä tiedonpoimintamenetelmässä.

Mima ja Ananiadou (2000: 186–187) päättelivät, että termit voidaan tunnistaa termiä määrittävän sanan avulla. Näistä termin läheisyydessä esiintyvistä sanoista Mima ja Ananiadou käyttivät nimitystä termikontekstisanat. C-painomenetelmä ja termikontekstisanojen käyttö yhdistyvät Miman ja Ananiadoun (2000) kehittämässä NC-painomenetelmässä, joka on tarkoitettu monisanaisten termien poimintaan. Tekstikontekstia hyödynnetään siis jo ainakin joissakin terminpoimintamenetelmissä.

Terminpoimintaohjelmat eroavat siinä, kuinka pitkiä lausekkeita ne poimivat. Tilastollisilla menetelmillä poimitaan lyhyitä jaksoja ja poiminta aloitetaan yleensä yhden sanan mittaisilla jaksoilla. Lingvistisillä menetelmillä poimittavien jaksoiden pituus vaihtelee yleensä kahdesta sanasta jopa kokonaiseen lauseeseen. Pitkien jaksoiden poimintaa perustellaan sillä, että terminologin on helpompi karsia hakutuloksesta ylimääräinen kieliaines kuin hakea puuttuva aines lähdekorpukselta.

Nimenomaan terminpoimintaan kehitettyjen terminologisten tai tietokone-lingvistiksen menetelmien lisäksi terminpoimintaan voidaan soveltaa korpuslingvistiikan menetelmiä. Lähtökohtana on oletus, että tarkasteltavan erikoisalan termejä ovat todennäköisesti sanat ja sanaliitot, jotka esiintyvät usein erikoisalan korpuksessa, mutta harvoin yleiskielen tai toisen erikoisalan korpuksessa. Yksinkertainen periaate on laskea tiettyjen leksikaalisten yksiköiden suhteellinen frekvenssi erikoisalan korpuksessa ja verrata sitä saman leksikaalisen yksikön frekvenssiin yleiskielen korpuksessa. Menetelmän etu on, että se on kielestä riippumaton. Sen heikkous on korpusten tarve ja se, että ainakin osa tutkittavista leksikaalisista yksiköistä on tunnettava etukäteen. Korpuslingvististä menetelmää ovat käyttäneet esimerkiksi Ahmad ja Salway (1997), jotka tutkivat turvallisuuskieltä (language of safety), ja Ahmad (1994; ks. myös Ahmad & Rogers 2001: 744; Uchimoto ym. 2000: 235).

#### **4.1.2.4 Puoliautomaattisen terminpoiminnan nykytila**

Puoliautomaattisten terminpoimintamenetelmien etu on niiden nopeus ja kyky toteuttaa virheettää ne tehtävät, jotka ohjelmat on ohjelmoitu tekemään. Puoliautomaattiset menetelmät ovat käyttökelpoisia ja luotettavia, kun käsitellään suuria tekstimassoja, jotka on tallennettu sähköisesti. Yleensä korpusten aineistosta osa on saatavana valmiiksi sähköisenä ja jäljelle jäävä osa on siirrettävissä sähköiseen muotoon nopeasti kuvanlukijan avulla. Tietokoneiden muisti ja tiedonkäsittelyn nopeus riittävät nykyisin suurtenkin tekstimäärien käsittelyyn.

Kaikesta innostuksesta ja lupaavista tuloksista huolimatta tutkijat korostavat, että puoliautomaattiset terminologisen tiedon poimintamenetelmät ovat vain apukeinoja eikä niiden käyttö ole yksioikoista (ks. esim. Ahmad & Rogers 2001: 740). Puoliautomaattinen terminpoiminta vaatii tietokoneen ja terminpoimintaohjelman. Jotta ohjelmaa voi käyttää terminpoimintaan, analysoitavan tekstin on oltava sähköisesti tallennettuna tietokoneen muistissa. Lisäksi termintunnistus saattaa vaatia muutakin tekstin esivalmistelua joko käsin tai toisella tietokoneohjelmalla. Joissakin tapauksissa aineiston tallennusmuoto tai tekijänoikeudet saattavat estää puoliautomaattisten menetelmien käytön. Osa lähteistä on vain paperimuodossa, ja tekijänoikeudet saattavat estää niiden tallentamisen sähköisesti.

Puoliautomaattisten terminpoimintamenetelmien puute on toistaiseksi myös se, että poimintatulokseen sisältyy runsaasti kelvottomia termiehdokkaita ja kelvollisia termejä

jää poimimatta. Osa termeistä saattaa jäädä termiehdokaslistalta pois siitä syystä, että moni markkinoilla olevista terminpoimintaohjelmista poimii vain substantiiveja tai substantiivilausekkeita. Näin muihin sanaluokkiin kuuluvat termit eivät pääse termiehdokaslistalle. Jos taas ohjelmassa ei ole sanaluokkarajoitusta, termiehdokkaiden määrä kasvaa huomattavasti, mutta suurin osa ehdokkaista on roskaa eli termiehdokkaita, jotka eivät ole tarkasteltavan erikoisalan termejä. (Laakso 2003: 21.) Myös pelkästään substantiiveja ja substantiivilausekkeita poimivan terminpoimintaohjelman tuottamassa termiehdokaslistassa on roskaa, sillä poimintatuloksessa on runsaasti vain yleiskieleen kuuluvia substantiiveja. Termikaavoja tunnistuskeinona käyttävien ohjelmien puute on, että yleinen termi, joka ei noudata tavallista termikaavaa, jää huomaamatta. Bowker ja Pearson (2002: 169) antavat esimerkiksi termin *peer-to-peer design*, jonka termikaava on N P N N (substantiivi + prepositio + substantiivi + substantiivi). Lisäksi jotkin termikaavat voivat olla tyypillisiä vain jollekin alalle. Lingvististen terminpoimintatyökalujen käyttöä rajoittaa myös se, että ne on suunniteltu yleensä yhtä kieltä varten eikä niitä voi helposti muuntaa toimimaan monen kielen kanssa (Bowker & Pearson 2002: 169). Termikaavat vaihtelevat kielestä toiseen, koska termien kielipolliset ominaisuudet ovat kielikohtaisia.

Roskaa eli valetermejä<sup>1</sup> ei pidetä yhtä suurena ongelmana kuin puuttuvia termejä eli piilotermiä<sup>2</sup>, koska ylimääräiset termiehdokkaat on verrattain helppo poistaa luettelosta, kun taas poimimatta jääneiden termien haku manuaalisesti teksteistä on suuritöistä ja vie aikaa. Lingvistisiä menetelmiä käytettäessä piilotermien määrää voidaan vähentää lisäämällä hakuehtoihin harvinaisia termikaavoja, mutta tämä muutos lisää valetermien määrää niin paljon, että haitat kasvavat helposti hyötyä suuremmiksi. Puutteet eivät johdu ohjelmien tekemistä virhevalinnoista vaan ohjelmoinnin puutteista, mikä puolestaan johtuu pitkälti siitä, että kieltä on vaikea mallintaa tietokonekielelle. Varsinkin lyhyistä, alle 10 sivun lähtöteksteistä tehdyn puoliautomaattisen terminpoiminnan tulos saattaa jäädä huonoksi. Lyhyillä lähtöteksteillä käsinpoiminta onkin huomattavasti varmempi ja jopa nopeampi menetelmä kuin automaattinen terminpoiminta.

## 4.2 Käsitetiedon poimintamenetelmät

Edellisissä jaksoissa olen kuvannut lyhyesti nykyisin käytössä olevia terminpoimintamenetelmiä, joilla lähtötekstistä voidaan tuottaa termiehdokaslista. Pelkkä termiehdokkaiden lista saattaa riittää joihinkin tarkoituksiin. Jos tavoitteena on esimerkiksi selvittää, onko tarkasteltavassa tekstissä tietyn erikoisalan termejä, termiehdokkaiden lista kuvaa tekstin sisältöä riittävän tarkasti. Mutta jos tavoitteena on esimerkiksi tehdä erikoisalan käsiteanalyysia tai verrata kahden kielen käsitteitä, ei termiehdokkaiden lista yksinään ole kovin käyttökelpoinen, sillä poimintamenetelmästä riippuen listassa saattaa olla synonymiaa ja variaatiota sekä valetermejä eli termiehdokkaita, jotka eivät ole tarkasteltavan erikoisalan termejä. Käsiteanalyysissa kiinnostavaa on tieto käsitesuhteista ja käsitepiirteistä, koska sen avulla terminologi pystyy rakentamaan käsitejärjestelmän. Kahden kielen käsitteiden vertailu ei sekään ole mahdollista ennen kuin käsitteiden väliset suhteet ja käsitepiirteet on selvitetty kummassakin kielessä.

Käsitesuhteiden ja käsitepiirteiden poimintamenetelmät voidaan jaotella sen mukaan, haetaanko tietoa teksteistä vai sanastoista. Perinteisessä käsiteanalyysissa (piirreanalyysi) käsitteiden väliset suhteet ja käsitepiirteiden kuvaama käsitesisältö selvitetään

<sup>1</sup> Käytän uudistermiä *valetermi* englanninkielisessä terminologian kirjallisuudessa esiintyvän termin *noise* vastineena. Valetermi on termiehdokas, joka ei ole tarkasteltavan erikoisalan termi.

<sup>2</sup> Englanninkielisen termin *silence* vastineena käytän uudistermiä *piilotermi*, joka on termiehdokaslistasta puuttuva termi.

analysoimalla esimerkiksi sanakirjojen ja sanastojen määritelmiä. Jos tavoitteena on rakentaa erikoisalan käsitejärjestelmä tai vertailla kahden kielen käsitteitä, ei sanakirjojen ja sanastojen käyttö tiedon lähteenä ole aina mahdollista eikä edes toivottavaa, sillä sanastoissa ja sanakirjoissa ei ole uusimpia käsitteitä. Uudella tai nopeasti kehittyvällä alalla tiedon lähteenä on käytettävä tekstejä, joissa on uusin erikoisalatieto. Käsitetiedon poimintaan teksteistä ei kuitenkaan voi suoraan soveltaa sanastomääritelmäanalyysejä, koska teksteissä käsitetieto on ilmaistu eri tavalla kuin sanakirjoissa ja sanastoissa. Terminologisten määritelmien sijasta teksteissä olevat määritelmät ovat usein ”vajaita” määritelmiä, joissa on joko yläkäsite tai erottava piirre mutta ei kumpaakaan. Lisäksi tekstien määritelmät saattavat olla tietyn asiantuntijan subjektiivisia näkemyksiä asiasta. Käsitteen määritelmäksi ei siis riitä yhden tekstin määritelmä, vaan määritelmän tulee olla globaali eli kaikkien käsitteen määritelmien yhdistelmä (Selov 2001: 8). Jos globaali määritelmä halutaan muokata tekstiaineiston pohjalta, on aineistoa oltava runsaasti ja eri kirjoittajilta. Tällöin ei aineiston käsittely perinteisin menetelmin enää ole mahdollista, vaan on käytettävä puoliautomaattisia käsitetiedon poimintamenetelmiä.

Viime vuosisadan lopulla tutkijat kiinnostuivat kielellisistä keinoista, joilla terminologisesti kiinnostava tieto voitaisiin paikallistaa teksteistä. Havaittiin, että teksteissä on kielellisiä ilmauksia tai graafisia merkkejä, joita voidaan käyttää apuna tällaisen tiedon poiminnassa. Esimerkiksi Sierra ja McNaught (2000: 10) mainitsevat, että on merkillepantavaa, miten käsitteiden esittely teksteissä noudattaa säännöllisiä kaavoja. Kirjoittaja voi muun muassa käyttää tiettyjä fraaseja (esimerkiksi *määrittelen*) ilmaistakseen eksplisiittisesti, että määritelmä seuraa (Pearson 1998: 103).

Tutkijat ovatkin päätyneet tulokseen, että tiettyjen kielellisten ilmausten tai graafisten merkkien avulla tekstistä voidaan kerätä tietoa termeistä, käsitesuhteista ja käsitepiirteistä (ks. esim. Bowker & Pearson 2002; Grinstead 2000; Kavanagh 1995; Pearson 1998). Nämä ilmaukset tai merkit voivat olla sanoja, sanaliittoja, lyhenteitä, välimerkkejä tai painoteknisiä keinoja. Piotrovskij, Rahubo ja Hažinskaja sanovatkin (1981: 36), että mikä tahansa sana, sanaliitto tai välimerkki voi toimia termin tai käsitetiedon ilmaisimena. Termin jäljessä voi esimerkiksi olla sulkeissa käsitetietoa, jolloin sulkeet ovat tutkijoiden hakema ilmaisun. Tutkijat pyrkivät toisin sanoen tunnistamaan ne kielelliset ilmaukset tai merkit, jotka osoittavat runsaasti käsitetietoa sisältävät kontekstit. Englanninkielisissä lähteissä näistä ilmauksista tai merkeistä käytetään usein termiä *knowledge probe*. Knowledge probe on kielellinen ilmaisu tai graafinen merkki, joka ilmaisee, miten tekstin kirjoittaja on hahmottanut käsitteiden väliset suhteet ja käsitepiirteet. Se on siis mentaalisen linkin ilmentymä tekstissä. Ilmausten avulla teksteistä voidaan tutkia esimerkiksi synonymiaa, ylä- ja alakäsitteiden välisiä suhteita (hyponymia ja hyperonymia), koostumussuhteita (meronymia) ja funktiosuhteita (Kavanagh 1995). Ahmad ja Rogers (2001: 749) lisäävät tähän luetteloon vielä antonymian tutkimisen.

Termin *knowledge probe* ottivat ensimmäisinä käyttöön Khurshid Ahmad ja Heather Fulford. He esittelivät vuonna 1992 menetelmän, jonka mukaan tiettyjä kielellisiä vihjeitä (linguistic cues) voidaan käyttää apuna termejä ympäröivän tekstikontekstin tutkimisessa, kun tekstistä etsitään käsitetietoa. Käsitettä selventääkseni annan kaksi esimerkkiä, jotka on lainattu Kavanaghin (1995) tutkielmasta:

- (1) Windows NT is a 32-bit operating system. (Kavanagh 1995: 25)
- (2) The mechanism includes three springs and a motor. (Kavanagh 1995: 27).

Esimerkissä 1 verbin ja artikkelin yhdistelmä *is a* ilmaisee ylä- ja alakäsitteen välistä hierarkkista geneeristä suhdetta, sillä Windows NT on eräs 32-bittinen käyttöjärjestelmä. Esimerkissä 2 verbimuoto *includes* ilmaisee koostumussuhdetta, jossa on yläkäsite *mechanism* (mekanismi) ja alakäsitteet *spring* (jousi) ja *motor* (moottori).

Tutkijat käyttävät käsitetietoa ilmaisevista kielellisistä ilmauksista ja merkeistä vaihtelevia nimityksiä, kuten *hinges* (Pearson 1996: 818–820), *linguistic signal* (Pearson 1998), *linguistic markers/indices* (Feliu & Cabré 2002: 2), *релятор* (Piotrovskij ym. 1981: 35), *текстовый маркер* (Šelov 1998: 161) ja *Меркmal* (Nuopponen 1993: 107). Käytän termin suomenkielisenä vastineena sanaliittoa *terminologinen koetin* tai tämän elliptistä muotoa *koetin*, koska ilmaisut tai merkit ovat tavallaan tiedon etsinnän työkaluja, aivan kuten kaivosalalla koettimet ovat malminetsinnän työkaluja. Näkökulma on siis tekstin lukijan, joka käyttää ilmauksia kuin malminetsijä koetinta tutkiakseen, onko ”maaperässä” eli tekstissä toivottuja aineksia. Käsitteelle *knowledge probe* on käytetty myös vastineita *ilmaisin* ja *merkitsin* (Soininen 1998: 17), mutta näistä sanoista syntyy se vaikutelma, että ne ovat tekstissä olemassa lukijasta riippumatta ja osoittavat tietynlaista ainesta ilman lukijan aktiivista panosta. Nimityksen taustalla on siis näkökulmaero. Nimitykset *ilmaisin* ja *merkitsin* ovat harhaanjohtavia myös siksi, että vain osa näistä kielellisistä ilmauksista ilmaisee terminologisesti kiinnostavaa tietoa. Esimerkiksi vain osa väitelauseista, joissa on *olla*-verbi muodossa *on* tai *ovat*, ilmaisee ylä- ja alakäsitteen välistä geneeristä suhdetta. Esimerkiksi lauseessa *Kanit ovat pehmeitä* kani ei ole pehmeän alakäsite. Ei ole myöskään selvyyttä siitä, esiintyykö termin tai määritelmän yhteydessä aina koetin. Muun muassa Soininen sanoo (1999a: 51), että teoriassa kielellisten rakenteiden käyttäminen tiedon poiminnassa on houkuttelevaa, mutta käytännössä yksiselitteisten rakenteiden määrittelemine on vaikeaa. Ongelma on samantapainen kuin termintunnistuksessa käytettyjen termikaavojen ongelma: paljon mahdollisia rakenteita jää tunnistamatta ja huomattava osa poimintatuloksesta on roskaa. Toistaiseksi terminologisia koettimia käsitelleet tutkimukset ovat paljolti keskittyneet englanninkielisiin aineistoihin, joita ovat tutkimuksissaan käyttäneet esimerkiksi Kavanagh (1995), Ahmad ja Rogers (2001), Bowker ja Pearson (2002), Condamines ja Rebeyrolle (1998), Grinstead (2000), Hearst (1992), Pearson (1996; 1998), Sierra ja McNaught (2000) sekä Vivaldi ja Rodriguez (2001). Jaksossa 7.6, jossa esittelen tutkimukseni suomalaisia ja venäläisiä koettimia koskevan empiirisen osuuden tulokset, esittelen myös joitakin Kavanaghin, Grinsteadin sekä Bowkerin ja Pearsonin tutkimia englanninkielisiä koettimia.

Menetelmää on kokeiltu myös ranskankielisillä aineistoilla. Soininen (1999a: 50–51) viittaa ranskankieliseen korpuspohjaiseen terminologiseen tietämyskantaan CTKB:hen, jonka luominen aloitettiin tunnistamalla termiehdokkaat ja jonka rakentamista jatkettiin etsimällä käsitesuhteet lausekaavojen avulla. Menetelmässä oletetaan, että termi esiintyy jonkin käsitesuhdetta kuvaavan terminologisen koettimen yhteydessä. Englannin- ja ranskankielisten aineistojen lisäksi tutkimusta on tehty ainakin tanskan-, espanjan- ja saksankielisellä aineistolla (Christensen 2000; Feliu & Cabré 2002; Nuopponen 1993).

Venäjän- ja ranskankielisellä aineistolla koetintutkimusta ovat tehneet jo 1980-luvun alussa Antonova (1984) sekä Piotrovskij ym. (1981). Venäläisistä tutkijoista aiheetta ovat myöhemmin käsitelleet ainakin Šelov (1998) ja Šilova (2005), jonka aineistossa oli venäjän- ja englanninkielisiä koneenrakennusalan tekstejä. Suomenkielisestä aineistosta ei koettimien toimintaa tietääkseni ole tutkittu. Erikielisillä teksteillä tehtyjen tutkimusten perusteella menetelmää on pidetty lupaavana, joten olen kokeillut sitä omaan suomen- ja venäjänkieliseen aineistooni. Koettimien tutkimuksessa käyttämistäni menetelmistä ja tutkimuksen tuottamista tuloksista kerron luvussa 7.

Työni alussa pohdin tutkimukseni kannalta tärkeitä terminologian teoreettisia kysymyksiä, kuten esimerkiksi *termin* käsitettä. Ennen kuin siirryn kuvaamaan työni empiiristä osuutta, jossa poimin teksteistä termiehdokkaita ja muuta käsitetietoa, palaan vielä *termin* käsitteeseen, jota tarkastelen nyt erityisesti terminpoiminnan näkökulmasta.



## 5 TERMI TERMINPOIMINNAN NÄKÖKULMASTA

Terminpoiminta on edelleen vuosikymmenten tutkimustyöstä huolimatta terminologinen ongelma. Jos termillä olisi tietyt muotopiirteet, joiden avulla se erotettaisiin muista sanoista, terminpoiminnan ongelmat olisi jo ratkaistu ja tietokoneohjelmat tunnistaisivat kaikki termit ja vain ne erehtymättä. Näin ei kuitenkaan ole. Muun muassa Pazienza (1998/1999: 197) päätyy toteamaan erästä termintunnistusmenetelmää käsittelevässä artikkelissaan, että menetelmän testaus on hankalaa, koska käsitykset termistä ovat epämääräisiä ja subjektiivisia (vrt. jakso 2.3). Terminmuodostus on kiinteässä yhteydessä käsitteenmuodostukseen, joka on kognitiivinen prosessi. Tämän prosessin kautta termeihin heijastuvat esimerkiksi käsitteen abstraktiotaso ja elinkaaren vaihe.

Termiyn arvioimiseksi on esitetty erilaisia menetelmiä. Sewangi (2001: 24–26) luettelee viisi keinoa, joilla termit on pyritty erottamaan muista sanoista:

- 1) rakenteelliset piirteet
- 2) frekvenssi
- 3) standardisointi
- 4) outous
- 5) erikoisalatieto.

Rakenteellisilla piirteillä termiä arvioitaessa termiehdokkaita tarkastellaan lingvistikseen näkökulmasta, joten rakenteellisia piirteitä voi nimittää myös lingvistisiksi piirteiksi. Tällaisia piirteitä ovat termin pituus, osien sanaluokka sekä morfologinen tai syntaktinen rakenne ja rakenteen pysyvyys. Näitä termin lingvistisiä piirteitä käsittelemme tarkemmin jaksoissa 5.2 ja 5.3. Lingvistiksi piirteinä voidaan pitää myös termivariaatiota ja siihen läheisesti liittyvää synonymiaa sekä ellipsiä. Synonymian ja termivariaation esiintymiseen vaikuttavat osittain kielenulkoiset syyt, kuten näkökulmaerot sekä käsitteiden ja termien elinkaaren vaihe. Käsittelemme synonymiaa ja termivariaatiota kuitenkin lingvististen tekijöiden yhteydessä jaksossa 5.4, koska ne ainakin osaksi liittyvät myös kielen rakenteeseen.

Lingvististen ominaisuuksien lisäksi termiehdokkaan termiä voidaan arvioida määrällisillä ominaisuuksilla, kuten Sewangin mainitsemalla frekvenssillä tietyssä tekstissä tai tekstikokoelmassa, erikoisalan teksteissä yleensä tai verrattuna frekvenssiin yleiskielen teksteissä. Termien frekvenssiä eli esiintymien määrää tietyssä tekstimassassa selostan jaksossa 5.1. Frekvenssiä käytetään yleisesti puoliautomaattisessa termin-tunnistuksessa, koska se on helposti mitattavissa ja tunnistettavissa teksteistä automaattisesti.

Lingvististen piirteiden ja frekvenssin lisäksi Sewangi mainitsee termiyn arvioinnin mittareiksi standardisoinnin, outouden ja erikoisalatiedon. Standardisointia on pidetty termiyn kriteerinä kielenhuoltoon suuntautuneissa terminologisissa koulukunnissa. Tiukan linjan edustajien mukaan sana tai sanaliitto muuttuu termiksi, kun se on virallisesti hyväksytty termiksi ja otettu standardiin. Standardisointi saattaa sopia termiyn mittariksi joissakin tutkimuksissa, mutta tämän tutkimuksen aineistosta poimittujen termiehdokkaiden arviointiperusteeksi se ei sovi, koska merenkulun turvallisuuskäsitteitä ei ole standardisoitu. En käsittele standardisointia tarkemmin, mutta totean kuitenkin, että jos pitäydytään tiukasti periaatteessa, jonka mukaan termejä on vain standardeissa, ajaudutaan noidankehään. Uutta käsitettä nimeävät nimitykset voisivat saada termin statuksen vasta vuosien kuluttua, vaikka ne olisivat erikoisalalla jo laajassa käytössä. Samalla suuri osa erikoisalalla vakiintuneista nimityksistä ei saisi termin statusta ollenkaan, sillä kaikki erikoisalan termit eivät koskaan päädy standardiin. Standardisointi koskee loppujen lopuksi vain marginaalista osaa eri erikoisalojen termistöistä. Tästä syystä todellista erikoisalan kielenkäyttöä ei voi tutkia standardisoinnin näkökulmasta



eikä termeiksi voi hyväksyä vain ihannetermin kriteerit täyttäviä termejä (vrt. jakso 2.3.1 ja 2.3.2).

Neljäs Sewangin mainitsema termiyden arvioimiskeino on tuttuus- ja outousperiaate. Outousperiaatteen käyttöä termiyden mittarina tukee Šelovin (2001: 11) tutkimustulos, jonka mukaan oudot sanat ovat todennäköisiä termejä. Šelov testasi intuitioon perustuvia hypoteeseja standardien ja sanakirjojen aineistoilla. Hyvin usein outous kielessä tarkoittaa samaa kuin vierasperäisyys. Aina ei vierasperäisyyttä ole helppo havaita, sillä kielet ovat täynnä lainoja. Urizar ym. (2000: 377) sanovatkin, että termi tai sen osa on lainasana hyvin suurella todennäköisyydellä. Sewangin mielestä outousperiaate on kuitenkin huono mittari, koska tuttuus tai outous on hyvin yksilöllistä ja riippuu terminpoimijan taustatiedoista. Sama sana voi myös olla sekä termi että yleiskielen sana, jolloin se on näennäisesti tuttu. (Sewangi 2001: 25.) Myöskään Pearsonin (1998: 19) mielestä tuttuus ei ole käypä mittari, koska sitä on mahdoton mitata objektiivisesti. Koska termin tuttuus tai outous riippuu terminpoimijan erikoisalatuntemuksesta, sivuan tätä arviointikeinoa lyhyesti manuaalisen terminpoiminnan käsittelyn yhteydessä jaksossa 6.6.

Viides Sewangin mainitsema termiyden arviointikeino on erikoisalatieto. Siihen on kiinnittänyt huomiota myös Karihalme (1996: 81), joka esittää väitöskirjassaan kuusi lekseemien erikoiskielisyyden tai termiyden arvioinnin perustetta:

- 1) informaation relevanttisuus erikoisalan asiaintilojen kannalta
- 2) tiedon kuuluminen erikoisalan diskurssin teemoihin
- 3) tarkoitteen kuuluminen erikoisalan kontekstiin
- 4) käsitteen kuuluminen erikoisalan käsitteistöön
- 5) käsitteen määriteltävyys erikoisalan kontekstissa
- 6) kielen ilmauksen termipiirteet.

Karihalmeen luettelemista termiyden arviointiperusteista vain yksi (kohta 6) on kielellinen, kaikki muut liittyvät erikoisalakontekstiin. Sekä Sewangin että Karihalmeen luetteloista käy ilmi, että kielelliset perusteet eivät riitä termiyden arvioinnissa. Arvioinnissa tarvitaankin henkilöitä, joilla on erikoisalatietoa, eli luotettavimmin arvioinnin tekevät erikoisalan asiantuntijat erikoisalakontekstin perusteella. Erikoisalatiedon ja terminologian tuntemuksen vaikutusta manuaaliseen terminpoimintaan arvioin jaksossa 6.6, jossa etsin vastausta kysymykseen, onko erikoisalatiedolla vaikutusta terminpoiminnan onnistumiseen.

Sewangin esittämästä termintunnistuskkeinojen luettelosta puuttuu yksi termiyden mittari, joka on monen tutkijan mielestä jopa termiyden edellytys. Tämä on määritelmän olemassaolo. Vaikka standardissa ISO 1087-1:2000(E/F) termin määritelmästä onkin jätetty pois vaatimus, että käsitteen tulee olla määriteltä, vastakkaisiakin näkemyksiä on. Moni tutkija on sitä mieltä, että termillä tulee olla selkeästi määriteltä merkitys.<sup>1</sup> Esimerkiksi Golovin ja Kobrin (1987: 61), Griněv (1993: 29–30), Gorodetskaja (1993: 30) ja Pearson (1998: 10) ovat sitä mieltä, että määritelmä on termiyden ehdoton tunnusmerkki. Golovin ja Kobrin toteavat, että ilmaus, jolla ei ole määritelmää, ei ole termi:

« -- дефинированность оказывается обязательным признаком термина: имеющий определение – термин, не имеющий определение – не термин.» (Golovin ja Kobrin 1987: 61)

<sup>1</sup> Suomalaisen terminologian tulkinnan mukaan termejä ei määritellä, vaan määritelmä liittyy aina käsitteeseen. Tässä tutkimuksessa noudatan tätä tulkintaa.

Šelov (2001: 6–7) pohtii kysymystä, liittyykö termiin aina välttämättä määritelmä, ja vastaa, että määritelmä on kyllä riittävä, mutta ei välttämätön ehto, koska jotkut termit ovat itsessään motivoituja tai käsitteen sisältö selviää lähitermeistä. Mielestäni määritelmä ei voikaan olla välttämätön, sillä kuten aiemmin jaksossa 2.3.1 totesin, kaikkia erikoisalan käsitteitä ei ole määritelty tarkoin ja yksiselitteisesti, vaan joidenkin käsitteiden rajat ovat hämäriä tai käsitesisällöstä ollaan erimielisiä. Termiyn riippuvuus määritelmän olemassaolosta onkin kohtuuton vaatimus ainakin terminologin tai kääntäjän näkökulmasta, sillä mistäpä nämä ei-asiantuntijat voisivat tietää, onko jokin käsite määritelty joskus jossakin, ellei määritelmää ole annettu siinä tekstissä, josta he termejä poimivat. Tutkimuksessani en käsittele määritelmän liittymistä termiin terminpoiminnan yhteydessä. Sen sijaan yhteenkuuluvuudella on merkitystä tutkimuksen toisessa osassa, jossa tutkin suomalaisia ja venäläisiä terminologisia koettimia (ks. jaksot 7.4.3 ja 7.5.3).

Ennen kuin siirryn selostamaan konkreettista terminpoimintaa, tarkastelen lähemmin joitakin terminpoimintaan vaikuttavia termin ominaisuuksia. Aloitan tarkastelun termiesiintymien frekvenssistä eli esiintymien määrästä tietyssä tekstissä tai tekstikokoelmassa. Tämän jälkeen pohdin terminpoimintaan vaikuttavia termien kielelliseen rakentamiseen liittyviä morfologisia ja syntaktisia piirteitä, kuten termin pituutta sanoina mitattuna, termin tai termin osien sanaluokkaa, synonymiaa, variaatiota ja ellipsiä. Lisäksi käsittelem termiin vaikuttavista pragmaattisista tekijöistä kontekstin vaikutusta.

## 5.1 Termiesiintymien frekvenssi

Kuinka monta kertaa termiehdokkaan pitää esiintyä tekstissä tai teksteissä, jotta se voi saada termin statuksen? Tämä kysymys on keskeinen varsinkin puoliautomaattisten termintunnistusmenetelmien kehitystyössä. Yleensä oletetaan, että termin frekvenssi on suurempi kuin yksi, koska tämä oletus parantaa oleellisesti puoliautomaattisten termintunnistusmenetelmien tulosta (esim. Justeson & Katz 1995 lähteessä Soininen 1998: 4). Termin frekvenssiä käytetään esimerkiksi terminpoimintamenetelmässä, jonka soveltamiseen tarvitaan tarkasteltavan erikoisalan korpuksen lisäksi jonkin toisen erikoisalan tai yleiskielen korpus. Tässä terminpoimintamenetelmässä termin frekvenssin käyttö termiyn kriteerinä perustuu Kageuran ja Uminon (1996: 280) esittämille oletuksille, joiden mukaan kielellinen yksikkö on todennäköisesti tietyn erikoisalan termi, jos

- 1) se esiintyy tarkasteltavalla erikoisalalla usein
- 2) se esiintyy vain tarkasteltavalla erikoisalalla
- 3) se esiintyy suhteessa useammin tarkasteltavalla erikoisalalla kuin yleiskielessä
- 4) sen frekvenssi on jollain lailla poikkeava (biased) erikoisalalla.

Termien frekvenssien laskemisesta on kuitenkin hyötyä terminpoiminnassa silloinkin, kun käytössä on vain yhden erikoisalan korpus. Tällöin oletetaan, että termiys on vahva, jos moni kirjoittaja käyttää termiä aiheeltaan samanlaisissa teksteissä tai jos termiä käytetään monta kertaa samassa tekstissä (Pekarskaja 1984: 12–13). Asiakirjan koolla ei välttämättä ole kovin suurta vaikutusta frekvenssiin. Katzin (1996: 16) tutkimus nimittäin todistaa, että jos sana esiintyy 200-sivuisen kirjan ja 20-sivuisen artikkelin ensimmäisellä sivulla, sana ei välttämättä esiinny kirjassa 10 kertaa useammin kuin artikkelissa.

Tutkimusten perusteella tiedetään myös, että yleiskielen yleisimmät sanat esiintyvät tasaisesti kaikissa teksteissä ja läpi koko tekstin, kun taas termit esiintyvät purskeina jopa yhden tekstin tasolla (ks. Lahtinen 2000: 78). Tähän tulokseen on päädytty, kun teksti on jaettu osiin ja tutkittu termien frekvenssejä kussakin osassa erikseen. Jonkin sanan tai sanaliiton poikkeava frekvenssi on siis sen termiyn tunnusmerkki tarkastel-

tavassa kontekstissa. Poikkeuksia ovat erikoisalan keskeiset termit, kuten esimerkiksi termi *alus*, joka esiintyy yleisesti ja tasaisesti merenkulun teksteissä. Luk’janenkovin ja Sergeevan (1978) kokoamassa englanninkielisten merenkulun termien frekvenssisanastossa aluksen vastine *vessel* onkin toiseksi yleisin merenkulun englannin termi. Tietoa purskeisuudesta käytetään hyväksi, kun ratkaistaan, onko termi tietyn erikoisalan termi, usealle tieteenalalle yhteinen termi vai tieteen termi yleensä. Kaikille tieteen aloille yhteisiä termejä ovat esimerkiksi *analyysi*, *menetelmä*, *prosessi* ja *järjestelmä*. (Gerd 1986: 38–39.) Näitä niin sanottuja luokkasanoja käytetään muun muassa yläkäsitteinä määritelmässä, joissa määriteltävällä käsitteellä on hyvin korkea abstraktiotaso (ks. Pearson 1996: 819).

Sanaliittotermien haussa esiintymien määrä on hyvä suunnannäyttäjä, koska se paljastaa, mitkä sanat esiintyvät aina tai lähes aina yhdessä (vrt. Soininen 1999b: 8). Pekarskajan (1984: 12–13) mielestä termiehdokkaan termiys kasvaa, jos sitä käytetään sanaliittotermien osana. 1–3 sanan mittaiset termit esiintyvät Pekarskajan mukaan jopa useammin pitemmän termin osana kuin yksinään. Kageuran ja Uminon (1996: 281) mukaan monisanaisen termin ja sen osien termiuden välillä voi olla kolmenlaisia suhteita:

- 1) monisanaisen termin termiudella ei ole tekemistä osien termiuden kanssa
- 2) monisanaisen termin termiudella on osittain tekemistä osien termiuden kanssa
- 3) monisanaisen termin termiudella ei ole tekemistä osien termiuden kanssa, mutta termiys korreloi positiivisesti pysyvyyden kanssa.

Näitä suhteita käytetään hyväksi terminpoimintaohjelmissa, joissa poiminta aloitetaan yksisanaisista termeistä. Monisanaiset termit poimitaan yksisanaisten termien avulla olettaen, että yksisanaiset termit ovat hyvin usein monisanaisten termien osia.

Termin osien frekvensseihin liittyy läheisesti vaatimus termin kielellisen rakenteen pysyvyydestä. Sitä voidaan mitata esimerkiksi laskemalla todennäköisyys, jolla sanaliittotermien osat esiintyvät aina yhdessä (Gorodetskaja 1993: 32). Pysyvyyden mittaamisella pyritään vastaamaan kysymykseen, onko vierekkäisten sanojen yhdistelmä termi vai vapaa kielen yksikkö. Esimerkiksi kaikki adjektiivin ja substantiivin muodostamat sanaliitot eivät ole termejä. Joissakin termintunnistusohjelmissa termin tunnistus perustuu ainakin osittain monisanaisten termiehdokkaiden osien pysyvyyden mittaamiseen (ks. esim. Ahmad & Rogers 2001: 748; Dias ym. 2000: 339–350). Ohjelmat antavat termiehdokkaille sitä korkeammat pisteet mitä suurempi on todennäköisyys, että sen osat esiintyvät aina yhdessä. En ole tutkimuksessani kuitenkaan laskenut termiehdokkaiden osien pysyvyyttä, vaikka testattavissa terminpoimintaohjelmissa pysyvyys saattaa olla yhtenä tekijänä termiehdokkaiden pisteytyksessä.

Termiesiintymien frekvenssin on todettu olevan hyvä termiuden kriteeri, ja sitä käytetäänkin laajalti hyväksi puoliautomaattisissa terminpoimintaohjelmissa. Se ei kuitenkaan voi olla ainoa termiuden kriteeri, sillä tutkimusten mukaan moni termi esiintyy teksteissä harvoin (ks. esim. Pearson 1998: 123). Esimerkiksi Daille, Habert, Jacquemin ja Royauté (1996: 208) päätyivät omassa tutkimuksessaan tulokseen, jonka mukaan kaksisanaiset termit esiintyivät keskimäärin 2,84 kertaa, kolmisanaiset termit 1,52 ja vähintään nelisanaiset termit 1,02 kertaa. Jos pyritään mahdollisimman täydelliseen käsitejärjestelmään, harvinaisetkin termit ovat tärkeitä. Alhaisen frekvenssin syy voi olla liian pieni korpus tai se, että termistä käytetään varianttia tai lyhennettä. Tekstissä saatetaan viitata termin nimeämään käsitteeseen monta kertaa, vaikka käsitettä nimeävä termi esiintyisi tekstissä täydellisessä muodossaan vain kerran. Termi ei kuitenkaan ole vähemmän tärkeä, jos se esiintyy harvoin (Bowker & Pearson 2002: 166). Itse asiassa harvinaiset termit ovat ainakin kääntäjän kannalta hankalimpia, koska niitä ei ole sanakirjoissa. Muun muassa kääntäjille tarkoitettuja sanakirjoja tutkinut Igor Kudashev

(henkilökohtainen tiedonanto 21.3.2003) on sanonut, että harvoin esiintyvät termit ovat informatiivisimpia. Tämä ilmenee muun muassa siten, että harvoin esiintyvät termit määritellään mahdollisesti lähes jokaisen esiintymän yhteydessä. Harvoin esiintyviä termejä ei ole sanakirjoissa eikä sanastoissa, jolloin niiden nimeämää käsitettä koskeva tieto on vain teksteissä. Bowker ja Pearson (2002: 167) varoittavatkin asettamasta esiintymien lukumäärävaatimusta liian korkealle puoliautomaattisessa terminpoiminnassa, koska tällöin yhden käsitteen hieman eri muodoissa esiintyvät nimitykset eivät ehkä pääse termiehdokkaiden listalle.

Olen valinnut termin tai termiehdokkaan frekvenssin yhdeksi vertailukohteeksi terminpoimintakokeessani, koska frekvenssi on helposti ja yksiselitteisesti mitattavissa tutkimusaineistosta. Toinen valintaperuste on sen laaja käyttö termiyn kriteerinä terminpoimintaohjelmissa.

## 5.2 Termin pituus

Termiehdokkaan pituutta merkkeinä tai sanoina mitattuna voidaan pitää yhtenä termiyn vaikuttavana ominaisuutena. Termi voi olla sana, yhdyssana tai sanaliitto (Sanastotyön käsikirja 1988: 70) tai jopa lause (Danilenko 1971: 8). Lausermejä ovat esimerkiksi merenkulussa, urheilussa tai armeijassa käytettävät komennot. Lähinnä englanninkielisillä aineistoilla tehtyjen tutkimusten perusteella on todettu, että termi on tarkkuusvaatimuksen vuoksi usein monisanainen (ks. esim. Lahtinen 2000: 90). Englanninkieliset termit ovatkin tutkimusten mukaan keskimäärin kahden sanan mittaisia (Justeson & Katz 1995 lähteessä Urizar ym. 2000: 375; ks. myös Ivina 2003: 74; Katz 1996: 24). Daillen ym. (1996: 207) 600 englanninkielisen termin otoksessa ja Ivinan (2003: 74) englanninkielisessä riskirahoitusta käsittelevässä aineistossa hieman yli puolet termeistä oli kaksisanaisia. Yksisanaisten termien osuus jäi kummassakin aineistossa noin 20 prosenttiin. Termin osien lukumäärä on kuitenkin kielestä riippuva, sillä esimerkiksi Urizarin ym. (2000: 374) tutkimassa baskin kielen sanakirja-aineistossa yksisanaisten termien osuus oli 41,9 prosenttia. Myös suomen kielessä yksisanaisten termien osuus on todennäköisesti suurempi kuin englannin kielessä, koska suomessa termin tarkkuusvaatimus on johtanut yhdyssanatermien runsauteen. Englannista lainatut monisanaiset termitkin saavat suomen kielessä usein yhdyssanavastineen. Soininen (1998: 4–5) antaa esimerkin tietojenkäsittelytieteestä, jossa monisanainen englanninkielinen termi *central processing unit* on saanut suomessa vastineen *keskusyksikkö*.

Kielten väliset erot on otettava huomioon terminpoimintaohjelmissa. Esimerkiksi Justesonin ja Katzin (1995) kehittämä terminpoimintamenetelmä perustuu olettamukseen, että termi koostuu vähintään kahdesta sanasta, koska englannin kielessä monisanaiset termit ovat paitsi yleisimpiä myös semanttisesti tärkein teknisten tekstien elementti (Katz 1996: 24). Suomen- tai venäjänkielisten termien poimintamenetelmät eivät voi perustua samalle olettamukselle, koska suomen ja venäjän kielen morfologinen rakenne poikkeaa englannin kielestä.

Termin pituudelle ei periaatteessa ole ylärajaa, mutta yli nelisanaiset termit ovat harvinaisia useimmissa kielissä (Soininen 1999b: 8). Yleensä englanninkielisten termien enimmäispituutena pidetään kuutta tai jopa vain neljää sanaa (ks. esim. Pearson 1998: 125). Tällöinkin neljäntenä sanana on usein luokkasana, esimerkiksi *system* (järjestelmä) tai *method* (menetelmä). Daillen ym. (1996: 205) mukaan vähintään neljän sanan pituisten termien osuus on marginaalinen (5 %). Samansuuntaiseen tulokseen on päätynyt myös Ivina (2003: 74), jonka englanninkielisessä tutkimusaineistossa vähintään neljän sanan mittaisien termien osuus oli 5,8 prosenttia. Myös venäjän termien maksimipituutena voidaan pitää neljää sanaa, sillä ainakin Grinëvin (1993: 54) laskelmien mukaan

98–99 prosenttia englannin ja venäjän termeistä on korkeintaan neljän sanan mittaisia. Ehdottomana ylärajana Grinëv pitää seitsemää sanaa. Hänen mukaansa pitkät termit ovatkin usein termien yhdistelmiä eivätkä itsenäisiä termejä.

Termin enimmäispituus riippuu kielestä ja myös tavasta, jolla sanojen lukumäärä lasketaan. Yhdyssanoja suosivissa kielissä, kuten suomessa ja venäjässä, enimmäispituus on alhaisempi kuin sanaliittoja suosivissa kielissä, kuten esimerkiksi englannissa. Termin pituuteen sanoina mitattuna vaikuttaa myös se, lasketaanko prepositiot sanoiksi vai ei. Olen tutkimuksessani laskenut sanaksi välilyöntien väliin jäävän merkkijonon, jolloin prepositiot ovat itsenäisiä sanoja. Kielten välisistä eroista huolimatta seitsemää sanaa voidaan hyvin pitää termin enimmäispituutena, sillä yli seitsemän sanan mittaisen termien osuus termistöstä on häviävän pieni. Lisäksi näiden termien pysyvyys on alhainen, sillä näin pitkän termin muistaminen tuottaa vaikeuksia tavallisille kielenkäyttäjille.

Monisanaisten termien synnyn syy saattaa olla tarkkuusvaatimus, joka johtaa helposti pitkien määriteketjujen syntyyn, varsinkin silloin, kun käsite on uusi. Koska kielissä vaikuttaa pyrkimys taloudellisuuteen, pitkä termi lyhenee käsitteen käydessä tutuksi. Lyheneminen tapahtuu joko määritteitä karsimalla tai ottamalla käyttöön lyhenne. Merenkulun englannissa lyhenteiden käyttö termeinä on tavallista, ja englannin kielen kautta lyhenteet ovat siirtyneet myös merenkulun suomeen. Lyhenne on toki taloudellinen, mutta voidaan kysyä, miten hyvin käsitteen sisältö avautuu lyhennetermin avulla niille, joille ala ei ole tuttu. Lyhenteillä on kuitenkin erityisasema termien joukossa, kuten jaksoissa 6.3.1 ja 6.3.2 ilmenee.

Termin pituuteen liittyy kysymys, mistä termi alkaa ja mihin se päättyy. Tämä on ehkä yllättäen terminpoiminnan hankalimpia kysymyksiä, jos termit poimitaan teksteistä. Tekstejä aineistonaan käyttävän terminpoimijan on ratkaistava, onko sanaliitto termi, kaksi termiä vai ilmaus. Termirajoja on yritetty määrittää paikantamalla tekstistä rajapyykkejä, joiden väliin termi varmasti jää. Esimerkiksi Surreyn yliopistossa kehitetty terminpoimintaohjelma *System Quirk* poimii termiehdokkaiksi sanoja, jotka esiintyvät suurifrekvenssisten yleiskielen sanojen ja välimerkkien välissä (Ahmad 1994: 10). Termi ei varmasti jatku pisteen eikä pisteeseen rinnastettavien välimerkkien, kuten kysymysmerkin, huutomerkkin, kaksoispisteen tai puolipisteen, yli. Näitä tavallisempi välimerkki teksteissä on pilkku, joka myös olisi käyttökelpoinen rajamerkki, elleivät jotkut termit jatkuisi pilkun toiselle puolelle. Esimerkiksi Meriteiden sääntöjen 3 säännössä määriteltyjen käsitteiden *alus*, joka harjoittaa kalastusta ja *alus*, jonka ohjailukyky on rajoitettu nimityksissä pilkku jää termin osien väliin. Tietysti tällaiset termit ovat harvinaisia, mutta jos poimintatulokseen pyritään saamaan kaikki termit, on poimittava nekin, joissa pilkku on termin osien välissä.

Termin pituuteen vaikuttavat olennaisesti määriteosat, joiden kuuluminen termiin ei aina ole itsestään selvää. *Pieni*, *suuri*, *pitkä* ja *lyhyt* ovat vältettäviä yleiskielen sanoja, kun termitetään (Sanastotyön käsikirja 1988: 85), mutta esimerkiksi Meriteiden säännöissä sanaliitot *pieni alus* (6 sääntö) ja *hyvin suuri alus* (7 sääntö) ovat termejä, koska aluksia luokitellaan muun muassa koon perusteella. Samassa asiakirjassa on myös käsite *hyvä merimiestapa* (8 sääntö). Kuuluuko adjektiivimäärite *hyvä* termiin vai ei? Kognitiivisen lingvistiikan edustaja Elzbieta Tabakowska (henkilökohtainen tiedonanto 25.8.2004) esittää, että pääsanana määritteet eivät ole sattumanvaraisessa järjestyksessä, vaan oliokäsitteen pysyvää ominaisuutta ilmaiseva määrite on lähempänä pääsanaa kuin tulkinnanvaraiset ominaisuudet. Tabakowskan englanninkielisessä esimerkissä *an uncomfortable huge round oak conference table* on oliokäsite *conference table*, jolla on joukko määritteitä. Koska valmistusmateriaali ja muoto ovat tarkoitteen pysyviä ominaisuuksia, ne ovat pääsanana vieressä. Sen sijaan koko, jota esimerkissä ilmentää adjek-



tiivä *huge*, on suhteellinen ominaisuus ja adjektiivin *uncomfortable* ilmaisema käyttömukavuus on täysin tulkinnanvarainen. Vastaavia puhujan tulkinnasta riippuvia määritteitä ovat esimerkiksi *hyvä* ja *huono*, joiden kuuluminen termiin on siis kyseenalaista.

Termin frekvenssiä esiintymien lukumäärällä mitattuna tai pituutta sanaluvulla mitattuna ei voida pitää yksinomaisina termintunnistuksen lähtökohtina, mutta kun nämä terminpöimintaan vaikuttavat kvantitatiiviset ominaisuudet yhdistetään kvalitatiivisiin tekijöihin, niistä on tunnistuksessa merkittävää apua.

### 5.3 Termin sanaluokka

Tutkijoiden mielipiteet termin sanaluokasta vaihtelevat, ja samaa mieltä tutkijat ovat lähinnä siitä, että termit ovat suurimmaksi osaksi substantiiveja tai substantiivilausekkeita. Moni tutkija hyväksyykin termeiksi vain substantiivit ja substantiivilausekkeet. Esimerkiksi Citkina (1988: 119) sanoo, että moniosaiset termit ovat sanojen tai termielementtien yhdistelmiä, joita käytetään nimeämiseen ja joiden pääsana on substantiivi:

Составные термины – это сочетания слов (терминологических элементов) номинативного характера с именем существительным в качестве ядерного слова. (Citkina 1988: 119)

Substantiivien ja substantiivilausekkeiden suosiminen perustunee ainakin osittain kielitieteessä yleiseen käsitykseen, että substantiivit ja substantiivilausekkeet kantavat tietoa (Lahtinen 2000: 50; ks. myös Danilenko 1971: 53). Substantiivilausekkeita pidetään konkreettisina, asiallisina ja tarkkoina. Joidenkin laskelmien mukaan noin 90 prosenttia termeistä on kielestä riippumatta substantiivilausekkeita (ks. esim. Soininen 1999b: 8; vrt. Daille ym. 1996: 207; Urizar ym. 2000: 375). Grinëvin (1993: 28–29) mielestä lähes kaikki termit ovat substantiiveja, sillä verbitermit voidaan aina muuttaa substantiiveiksi ja adjektiivit tai adverbit ovat termin osia eivätkä itsenäisiä termejä (Grinëv 1993: 33). Substantiivien ja substantiivilausekkeiden suuri osuus on ymmärrettävä, sillä suurin osa käsitteistä on oliokäsitteitä (konkreettisia olioita, aineita, artefakteja, elollisia olioita jne.), joita nimetään kielellisesti substantiiveilla (L’Homme 2003: 404). Oliokäsitteiden lisäksi substantiivit voivat nimetä muidenkin käsiteloukkien käsitteitä (Komarova 1991: 85).

Puoliautomaattisessa terminpöiminnassa onkin yleensä keskitytty substantiivilausekkeiden hakuun. Terminpöimintaohjelmien algoritmin taustalla on ennako-oletus, että termi on yksi- tai monisanainen substantiivilauseke ja että substantiiviin päättyvät adjektiivien ja substantiivien yhdistelmät ovat todennäköisiä termejä. (Ks. esim. Katz 1996: 25; Pazienza 1998/1999: 184; Vivaldi & Rodríguez 2001: 35.) Justeson ja Katz tämentävät, että termiehdokas on monisanainen substantiivilauseke, jossa pääsana on substantiivi ja määriteosa on substantiivi tai adjektiivi. Lisäksi pääsanana ja määriteosan välissä voi olla prepositio. (Justeson & Katz 1995 lähteessä Gamper & Stock 1998/1999: 151.)

Substantiivit ja substantiivilausekkeet sopivat hyvin käsitteiden nimeämiseen, mutta Pilkkeen (2000: 66) mielestä verbitkin pitäisi hyväksyä termeiksi, koska ne tuovat esiin käsitteen dynaamisuuden. Nykyisin yhä useampi tutkija onkin valmis hyväksymään termeiksi myös verbit (Danilenko 1971: 8; L’Homme 2003: 403; Pearson 1998: 10; Tatarinov 1996: 200). Kollokaatioita tutkinut L’Homme (2000: 92, 96–99) sanoo, että substantiivien kanssa kollokaatioita muodostavat verbit voivat olla sellaisenaan termejä eivätkä vain termin osia. L’Hommen mukaan verbi voi olla termi, jos se on alalle tyy-

pillinen, sitä käytetään erikoisalalla muussa kuin sen tavallisessa yleiskielisessä merkityksessä tai monen substantiivin kanssa.

Verbitermit ovat tyypillisiä joillakin erikoisaloilla, kuten esimerkiksi lakikielessä (tulla voimaan) ja tietojenkäsittelytieteessä (tulostaa, päivittää) (Soininen 1998: 5; ks. myös Urizar ym. 2000: 375). Tuskin on erikoisalaa, jolla ei olisi ollenkaan verbitermejä, sillä verbit kuvaavat tyypillisesti prosesseja ja toimintaa ja (verbeistä johdetut) substantiivit prosessin tulosta. Ne, jotka eivät verbejä termien joukkoon hyväksy, perustelevat kantaansa sillä, että verbin voi aina nominalisoida. Tämä on ehkä totta ja perusteltua, jos hyväksytään ajatus, että verbistä johdettu substantiivi ja verbi ovat saman käsitteen nimityksiä, mutta onko järkevää nominalisoida esimerkiksi edellä mainittua verbiä *tulla voimaan*? Ei liene kielellisesti parempi sanoa ”Asetuksen voimaantulo tapahtuu 1. heinäkuuta 2002” kuin ”Asetus tulee voimaan 1. heinäkuuta 2002”? Koska suomen kielessä suositaan verbejä, tuntuu mahdottomalta ajatella, ettei verbejä hyväksyttäisi termeiksi. Venäjän kielessä sen sijaan suositaan verbien sijasta substantiiveja, mutta tästä huolimatta ainakin Danilenko oli jo 1970-luvulla valmis hyväksymään verbit termeiksi.

Suuri osa terminologeista on valmis hyväksymään termeiksi substantiivit, substantiivilausekkeet ja jopa verbit, mutta adjektiivien esiintyminen itsenäisinä termeinä on asetettu kyseenalaiseksi. Jotkut tutkijat ovat sitä mieltä, että adjektiivi ei yksinään voi olla termi, vaan se esiintyy aina substantiivin kanssa yhdessä, jolloin substantiivin ja adjektiivin muodostama sanaliitto on termi. Mutta esimerkiksi Danilenko (1971: 45) jakaa adjektiivit kahteen ryhmään: osa adjektiiveista on termejä ilman substantiivia (виртуальный, визуальный). Nämä adjektiivit voivat vapaasti muodostaa sanaliittoja substantiivien kanssa ja niiden merkitys säilyy muuttumattomana sanaliitossakin. Osa adjektiiveista sen sijaan vaatii substantiivin seurakseen. Adjektiivitermien osuus on alasidonnainen, ja esimerkiksi muotoilun alalla niiden osuus on merkittävä (Karihalme 1996: 31).

Adjektiivien osuus termeistä on varsin pieni, mutta sitäkin pienempi on adverbien osuus, sillä tutkijoiden mielestä adverbi voi yleensä olla termissä vain termin osana (Danilenko 1971: 56). Adverbeja on kuitenkin termeinä erityisesti musiikin kielessä, mutta myös merenkulussa, kuten esimerkiksi Danilenkon (1971: 49) mainitsevat venäläiset termit *наливом* ja *навалом* (irtona).

Terminpoiminnassa tietoa termien ja termin osien sanaluokista voidaan käyttää hyväksi termikaavojen muodossa. Esimerkiksi englannin kielessä yksi yleisimmistä terminmuodostustavoista varsinkin tekniikan aloilla on adjektiivin ja substantiivin tai kahden substantiivin yhdistäminen (L’Homme ym. 1996: 297; Bowker 1997: 276; Ivina 2003: 74; ks. myös Citkina 1988: 68). Tämän tiedon perusteella englanninkielinen termiehdokas, jonka termikaava on  $A_x N_x$ , on todennäköinen termi (Katz 1996: 25). Termikaavassa A on adjektiivi ja N on substantiivi. Alaindeksi  $_x$  osoittaa, että termissä voi olla monta adjektiivimääritettä ja myös substantiivimääritteitä, mutta viimeisenä termissä on aina substantiivi. Justeson ja Katz (1995) ilmaisevat saman asian kaavalla  $((A|N)+|((A|N)^*(NP)?)(A|N)^*)N$ .

Termin sanaluokka siis auttaa termintunnistuksessa, mutta jos tekstistä halutaan poimia kaikki termit ja vain termit, ei sanaluokka tai aiemmin mainittu termin frekvenssi ja pituus yksin tai edes yhdessä riitä, sillä kuten Pearson (1998: 10) asian ilmaisee, termeillä on oltava selkeästi määritelty merkitys, kun niitä käytetään siinä kontekstissa, jota varten ne on määritelty:

These [terms] may be nouns, verbs, adjectives or adverbs which are considered to have a clearly defined meaning when used in the context for which they have been defined.  
(Pearson 1998: 10)



Termin merkitys on kiinteästi kytköksissä käsitteeseen, jota se nimeää. Frekvenssin, pituuden ja kieliopillisen rakenteen lisäksi terminpoiminnassa onkin otettava huomioon synonymia, polysemia, termivariaatio ja ellipsin käyttö, joiden tunnistamisessa käsitetieto on välttämätöntä. Nämä tekijät hankaloittavat terminpoimintaa, sillä toisin kuin frekvenssien laskemisessa ja kieliopillisten rakenteiden tunnistamisessa, ei automaattisista terminpoimintamenetelmistä ole apua synonymian, polysemian tai termivariaation tunnistamisessa. Seuraavaksi käsittelen näitä yleisiä kielen ilmiöitä terminpoiminnan näkökulmasta.

#### 5.4 Saman käsitteen eri nimitykset

Perinteisen terminologian teorian tavoite on, että käsitettä nimeää yksi termi ja termi nimeää vain yhtä käsitettä (ks. jakso 3.1). Standardisointijärjestöt tekevät töitä tähän tavoitteeseen pääsemiseksi, mutta käytännössä erikoiskielen käsitettä usein nimetään eri nimityksillä eli synonymia on tavallista. Synonymia määritellään termien väliseksi suhteeksi, jossa termit nimeävät samaa käsitettä (ISO 1087-1: 2000(E/F): 8). Termin synonyymit kertovat muun muassa näkökulmaeroista, jolloin termissä voi korostua jokin käsitepiirre toisten kustannuksella. Esimerkiksi termillä *aluksen rakenteellinen turvallisuus* on synonyymi *aluksen passiivinen turvallisuus*, jossa korostuu ihmisen mahdollisuus tai mahdottomuus vaikuttaa turvallisuuteen silloin, kun alusta käytetään. Wüsterkin myönsi, että synonyymit ovat usein välttämättömiä, kun on tarpeen painottaa saman käsitteen eri puolia (Tatarinov 1996: 43). Synonyymisten nimitysten käyttö saattaa olla välttämätöntä myös liiallisen toiston välttämiseksi, varsinkin jos erikoisalan teksti käsittelee hyvin suppeaa alaa.

Jotkut tutkijat (esim. Koivisto-Alanko 2000: 56) ovat sitä mieltä, että täydellinen synonyymia on harvinaista, mutta ainakin lyhenteet hyväksytään yleisesti täydellisiksi synonyymeiksi. Synonyymi voi olla myös vieraskielinen termi, jota käytetään omakielisen termin rinnalla, koska vieraskielistä termiä pidetään neutraalina (Lotte 1982: 21, 40). Pearsonin (1998: 170) mukaan täydellisessä synonyymiassa synonyymit voidaan korvata toisillaan kaikissa konteksteissa tietyllä alalla. Osittaissynonyymiasta on kyse, jos synonyymit eivät nimeä täysin samaa käsitettä. Osittaiset synonyymit ovat tärkeitä, kun käsiteluoikkien rajoja käydään.

Synonyymien käyttöön on kielellisiä syitä, mutta usein syyt ovat sosiaalisia tai historiallisia. Sosiaalinen syy on edellä mainittu asiantuntijoiden näkemysero eli halu korostaa tiettyä käsitepiirrettä käsitettä nimeävässä termissä. Historiallinen syy on esimerkiksi lainasanojen käyttö. Synonyymien yleisyys riippuu erikoisalasta, kuten Pilkkeen (2000) tutkimus osoittaa. Pilkkeen tutkimissa tekniikan sanastoissa keskimäärin 15 prosentille termeistä oli sanastossa annettu ainakin yksi synonyymi, joka oli usein termin lyhennetty muoto. Lääketieteen sanastoissa synonyymeja oli paljon useammin, sanastosta riippuen jopa 51 prosentilla termeistä. Oikeustieteen sanastossa sen sijaan oli annettu synonyymi vain 5 prosentille termeistä. (Pilkke 2000: 281–284.)

Synonyymeista kiinnitän erityistä huomiota lyhenteisiin, koska suomen- ja venäjänkielisissä merenkulun teksteissä on runsaasti isokirjainlyhenteitä paljolti englannin kielen vaikutuksesta. Onkin tavallista, että monisanaisia termejä (complex terms) edustavia lyhenteitä esiintyy erikoisalojen teksteissä runsaasti (ks. esim. Fontenelle 2000: 351). Käsittelen lyhenteitä synonyymeinä, vaikka ne voidaan ajatella myös termin varianteiksi, kuten esimerkiksi Daille ym. (1996: 201) tekevät.

Synonymiaan läheisesti liittyvä terminologinen ilmiö on termivariaatio, joka ilmenee termeissä monella tavalla. de Bessé ym. (1997: 155) pitävät termivarianttina termiä, jolla on poikkeava kirjoitusasu. Suomen kielessä termivariaatio ilmenee yksinkertai-

simmillaan moniossaisten termien yhdyssana- tai sanaliittomuotoina. Esimerkiksi *meriliikenteen ohjausjärjestelmä* ja *meriliikenteenohjausjärjestelmä* ovat saman termin variantteja. Tällaisen variaation useimmat tutkijat sallivat termeille. de Bessén ym. mukaan myös termin supistetut ja laajennetut muodot ovat variantteja. Daille ym. (1996: 201) määrittelevät termin variantin lausumaksi, joka liittyy alkuperäiseen termiin semanttisesti ja käsitteellisesti. Variantti ei siis heidän mukaansa ole välttämättä täsmälleen saman käsitteen nimitys. Synonymian ja termivariaation ero on häilyvä, mutta periaatteessa varianteilla on alkuperäinen termi, jonka muunnelmia variantit ovat. Synonyymiset termit sen sijaan voivat syntyä ilman lähtötermiä. Esimerkiksi termit *kantarelli* ja *keltavahvero* nimeävät samaa sienilajia (*Cantharellus cibarius*), mutta niillä ei ole yhteistä suomenkielistä lähtötermiä. Tässä tutkimuksessa tarkoitan termivarianteilla de Bessén ym. tavoin saman käsitteen nimityksiä, jotka poikkeavat toisistaan kirjoitusasultaan. Tämän määritelmän mukaiset variantit voidaan korvata toisillaan kontekstista riippumatta (Gorbunov 2004: 14). Tekstissä ero näkyy siten, että termin variantteja ei yleensä kirjoiteta tekstiin peräkkäin, kun taas synonyymisten termien kanssa näin menettellään usein.

Termivariaatio voidaan ryhmitellä esimerkiksi syntaktiseen, morfosyntaktiseen ja semanttiseen variaatioon, kuten Jacquemin tekee (1997: 253–254; ks. myös Urizer ym. 2000: 379). Tässä ryhmittelyssä syntaktista variaatiota ovat rinnastukset (coordinations; *tropical and citrus fruits*), jossa kahden rinnasteisen termin yhteinen terminosa kirjoitetaan vain kerran, muunnokset (modifications; *bottom-up chart parser* ja *bottom-up parser*) ja permutaatiot (permutations; *alignment of sentence* ja *sentence alignment*). Rinnastukset ovat yleisiä kielestä riippumatta. Sen sijaan permutaatio, jossa termin osien järjestys vaihtelee, on englannin kielessä tavallinen termivariaation muoto (Kageura & Umino 1996: 274). Morfosyntaktisessa variaatiossa ilmenee syntaktisen variaation lisäksi vähintään yhden sanan morfologinen muunnos, esimerkiksi substantiivin ja verbin, kahden substantiivin tai substantiivin ja adjektiivin välillä (*acquisition of lexicon* ja *lexical acquisition*).

Semanttinen variaatio poikkeaa Jacquemin mukaan kahdesta muusta variaatiotyypistä. Siinä termit liittyvät toisiinsa semanttisesti, kuten esimerkiksi termit *formal interpretation* ja *formal semantics*, koska termi *interpretation* liittyy semanttisesti termiin *semantics*. Urizer ym. (2000: 379) sanovat, että semanttista variaatiota edustavat variantit, joissa jokin termin osa on korvattu synonyymillaan (*kidney function* ja *renal function*; *meriliikenteen ohjaus* ja *alusliikenteen ohjaus*). Daille ym. (1996: 200) lisäävät variaatioluokitteluun neljäntenä morfologisen variaation ilman samanaikaista syntaktista variaatiota. Oleellista Daillen ym. jaottelussa on se, säilyykö alkuperäinen termi ehjänä variaatiossa vai muuttuuko se lisäysten tai permutaation vuoksi. Alkuperäisen termin muuttuminen tuottaa ongelmia automaattiselle termintunnistukselle (Daille ym. 1996: 200).

Tässä yhteydessä on syytä huomauttaa, että myös ellipsi on eräs variaation muoto (Jacquemin 1997: 260). Elliptisessä termissä jokin termin osa on jätetty pois, jolloin yhdyssanan tai sanaliiton lyhyt muoto korvaa käsitteen nimityksen. Esimerkiksi *väylä* on *meriväylän* elliptinen muoto. Hierarkkisessa käsitejärjestelmässä elliptinen termi on usein yhtä tasoa ylemmän käsitteen nimitys, jota käytetään nimeämään alakäsitettä. Tekstiyhteys auttaa kuitenkin lukijaa tulkitsemaan lukemansa oikein. Tatarinov (1996: 187) käyttää ilmiöstä nimitystä hyperonyyminen variaatio (гипонимическое варьирование). Tutkimuksessani termivariaation käsite on ymmärretty laajasti niin, että sen alaan kuuluvat myös elliptiset termit.

Syntaktisen, morfosyntaktisen ja semanttisen variaation lisäksi termeille on tavallista ortografinen variaatio, jossa termivarianttien kirjoitusasussa on ehkä vain yhden kir-

jaimen ero (vrt. Ahmad & Rogers 2001: 754). Tätä variaatiotyyppiä ei Jacquemin ole edes maininnut, mutta mielestäni sekin on otettava huomioon, vaikka se ei terminologian kannalta ehkä ole yhtä merkittävää kuin muut variaatiotyypit. Suomen kielessä on hyvin tavallista, että moniosaisten termien osat kirjoitetaan milloin yhdyssanaksi, milloin erikseen, kuten esimerkiksi termi *meriliikenteen ohjausjärjestelmä*, joka saatetaan kirjoittaa myös yhteen *meriliikenteenohjausjärjestelmä*. Esimerkkitermissä kumpikin kirjoitusmuoto on terminologisesti samanarvoinen, mutta joskus sillä, kirjoitetaanko sanat yhteen vai erikseen, voi olla terminologista merkitystä. Salmi (2003: 12) esimerkiksi kysyy, ovatko sanaliitot *lämmön säätely*, *hallitusten välinen* tai *tiedon hallinta* termejä. Hän esittää nyrkkisäännöksi, että termejä nämä ovat yhteen kirjoitettuina, mutta jos puhutaan lämmön säätelystä tai tiedon hallinnasta yleensä, sanat kirjoitetaan erilleen. Osoittavatko erilaiset kirjoitusasut siis kirjoittajan epävarmuutta siitä, ovatko ilmaiset termejä vai eivät, vai osoittaako horjuvuus oikeinkirjoitussääntöjen tuntemuksen puutetta? Tähän kysymykseen ei tutkimukseni aineistosta löydy vastausta, mutta pidän kirjoitusasun vaihtelua ortografisena variaationa ja variantteja samanarvoisina termeinä.

Tutkimuksessani käsittelen ellipsiä, syntaktista, morfosyntaktista ja ortografista variaatiota sekä sellaista semanttista variaatiota, jossa termivariantit nimeävät samaa käsitettä. Noudatan tässä rajauksessa Montero-Martínezin ym. (2002: 190) näkemystä, jonka mukaan termivariantit nimeävät samaa käsitettä. Näiden variaation kielellisten ilmenismuotojen lisäksi variaatiotyyppien jaotteluperusteena voidaan käyttää aikaa, jolloin voidaan puhua synkronisesta ja diakronisesta termivariaatiosta. Synkronisella termivariaatiolla tarkoitan variaatiota, joka ilmenee hyvin lyhyen ajanjakson aikana saman asiakirjan sisällä tai suunnilleen samaan aikaan ilmestyneiden asiakirjojen välillä. Diakronisella termivariaatiolla tarkoitan variaatiota, joka ilmenee termien muuttumisena pitkän ajan kuluessa. Jotta diakronisen termivariaation voisi erottaa synkronisesta termivariaatiosta, täytyy tarkasteltavan ajanjakson olla riittävän pitkä, sillä kieli muuttuu hitaasti. Uusien käsitteiden nimitykset muuttuvat joskus nopeastikin. Esimerkiksi käsitteen *alusliikennepalvelu* nimitykset ovat muuttuneet melkoisesti noin kymmenen vuoden aikana. En käsittele diakronista variaatiota lähemmin tässä tutkimuksessa, mutta olen käsitellyt aihetta aiemmin julkaistussa artikkelissani (Pasanen 2003).

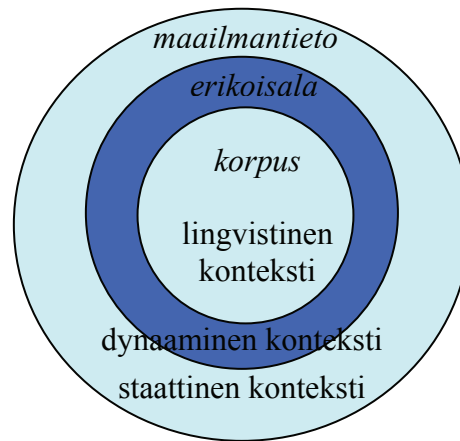
Diakronista terminologista tutkimusta on tähän mennessä tehty vähän, vaikka useissa tutkimuksissa on todettu, että se olisi hyvinkin tarpeellista. Esimerkiksi Tatarinov (1996: 136) toteaa, että diakroninen termivariaatio kuuluu tieteen kieleen. Suomenkielisistä terminologian alan tutkimuksista Karihalmeen tutkimus *Muotoilusanaston termistytminen* (1996) edustaa diakronista termintutkimusta puhtaimmillaan, vaikka Karihalmeen teoreettinen viitekehys ei olekaan terminologian teoria. Karihalme (1996: 278) päätyi toteamaan, että termivariaatio on yleistä jo yhden tekstin sisällä puhumattakaan eri kirjoittajien kirjoittamien ja eri aikoina kirjoitettujen tekstien välillä. Termivariaatiota ei siis voi sivuuttaa sen paremmin terminologisessa tutkimuksessa kuin sanastotyösäkään.

Synonymian ja variaation lisäksi myös ammattislangin ilmaus voi nimetä samaa käsitettä kuin termi. Ammattislangi eroaa termeistä ennen kaikkea siten, että ammattislangi ei ole normitettua. Sen sijaan se on tyyllillisesti puhekielistä ja värittyä (esimerkiksi *срапом* 'yliperämies') (Gerd 1986: 40–41; Kazarina 1998: 31). Vaikka ammattislangin värikkäät ilmaukset voivat olla esivaihe matkalla termiyeen, jää sinänsä kiinnostava merenkulun slangi tutkimukseni ulkopuolelle, koska aineistossani ei ole ammattislangin ilmauksia.

## 5.5 Konteksti termien ja käsitetiedon poiminnassa

Olen edellä tarkastellut termien kvantitatiivisia ja lingvistisiä ominaisuuksia, joita koskeva tieto on tarpeen termien tunnistamisen ja poiminnan automatisoinnissa. Lisäksi termien ja varsinkin käsitetiedon poimintaan vaikuttaa konteksti, sillä vaikka perinteisen terminologian teorian mukaan termin käyttö ei riipu kontekstista, käytäntö on osoittanut, että pelkän termiehdokaslistan perusteella on ainakin joistakin termiehdokkaista vaikea sanoa, mitä käsitettä ne nimeävät ja ovatko ne termejä ollenkaan (vrt. Antia 2000: 115–116). Määrälliset ja kielelliset mittarit, kuten termin pituus sanoina, frekvenssi ja termikaava, eivät siis riitä termintunnistuksessa, sillä termiymyden kannalta ratkaisevaa on muodon sijasta käyttö ja lopulta erikoisalakonteksti laajasti ymmärrettynä. Joissakin tapauksissa on tarkasteltava erikoisalakontekstiäkin laajempaa kontekstia, jota kutsun maailmantiedoksi.

Konteksti voidaan määritellä hyvin monella tavalla tavoitteesta ja näkökulmasta riippuen. Sanastotyön yhteydessä kontekstilla tarkoitetaan yleensä käsitteen nimityksen tekstiympäristöä. ISO-standardissa 1087-1 konteksti on määritelty tekstiksi, joka kuvaa käsitettä tai käsitteen nimityksen käyttöä (ISO 1087-1:2000(E/F): 12). Laajemmin ymmärrettynä se voidaan jakaa Heikkisen (2003: 3) tavoin lingvistiseen eli kielensisäiseen kontekstiin ja ekstralingvistiseen eli kielenulkoiseen kontekstiin. Myös Kostera (1996: 284) laajentaa kontekstin käsitteen kielen ulkopuolelle. Hänen mukaansa konteksti on teksti- tai muu ympäristö eli asia-, lause-, teksti- tai kulttuuriyhteys. Diskurssintutkimuksessa käytetään termejä *lause- ja episodikonteksti*, *vuorovaikutuskonteksti* ja *kulttuurinen konteksti* (Jokinen, Juhila & Suoninen 2004: 30–33). Vuorovaikutuskonteksti on tilanne, jossa diskurssi tapahtuu. Terminologiaan sovellettuna se on tilanne, jossa erikoisalaviestintä tapahtuu. Kulttuurinen konteksti vastaa maailmantietoa, joka pysyy samana tilanteesta riippumatta. Sekä vuorovaikutuskonteksti että kulttuurinen konteksti sijoittuvat Heikkisen mainitseman ekstralingvistisen kontekstin alueelle. Kognitiivisen lingvistiikan edustajat Croft ja Cruse (2004: 102–103) jakavat kontekstin neljään tasoon. Ensimmäisenä on kielellinen konteksti, johon kuuluu aiempi diskurssi, välitön lauseyhteys ja tekstilaji. He sisällyttävät kielelliseen kontekstiin myös erikoisalan. Toisena on fyysinen konteksti, jolla Croft ja Cruse tarkoittavat viestintätapahtumaan osallistuvien aistein havaittavaa ympäristöä. Kolmantena on sosiaalinen konteksti, joka viittaa viestintään osallistuvien tilanteeseen ja keskinäisiin suhteisiin. Neljäntenä on niin sanottu tallennetun tietämyksen (stored knowledge) konteksti, joka vastaa maailmantietoa. Fyysinen ja sosiaalinen konteksti voidaan yhdistää tekstinulkoiseksi erikoisalakontekstiksi, jossa viestintä tapahtuu. (Ks. myös Yli-Jokipii 2006: 109–110.) Tarkasteltavan erikoisalan näkökulmasta voidaan siis puhua kolmen tason kontekstista. Ulommainen, staattinen konteksti on erikoisalan ulkoista kontekstia eli maailmantietoa ja keskimmäinen, dynaaminen konteksti on tarkasteltavan tekstin ulkoista erikoisalakontekstia. Yhdessä nämä kontekstin tasot muodostavat ekstralingvistisen kontekstin. Kolmas kontekstin taso eli lingvistinen konteksti on tarkasteltavan tekstin tai korpuksen konteksti.



KUVIO 5. Kontekstin kolme tasoa terminologisen tiedon poiminnassa

Kun termejä ja käsitetietoa poimitaan teksteistä, kaikki kolme kontekstitasoa on otettava huomioon. Termiehdokaslistan termiehdokkaiden arvioinnissa on palattava tekstikontekstiin eli tutkittava termiehdokkaan ympäristöä tekstissä ja haettava sieltä käsitetietoa eli tietoa käsitteiden välisistä suhteista ja käsitepiirteistä. Termiä ympäröivä tekstikonteksti voi myös johtaa uusien termien jäljille semanttisten linkkien kautta (Vivaldi & Rodríguez 2001: 38). Myös Pearson (1998: 10, 36) pitää kontekstia oleellisesti termiin liittyvänä sanoessaan, että termejä voidaan pitää termeinä vain, kun niitä käytetään tietyssä kontekstissa. Tässä kontekstin käsitteen voi laajentaa tarkoittamaan tekstikontekstin lisäksi erikoisalakontekstia. Riggs (1993: 207) kysyykin, onko mahdollista tai edes toivottavaa ymmärtää termejä sen kontekstin ulkopuolella, jossa niitä käytetään. Riggsin mainitsema konteksti on tässä tulkittava erikoisalakontekstiksi. Erikoisalakonteksti luo kehyksen, joka auttaa terminpoimijaa ratkaisemaan, mitkä termiehdokkaat kuuluvat terminlistaan (Antia 2000: 115). Erikoisalakonteksti myös ohjaa terminpoimijaa etsimään termejä, jotka puuttuvat tutkittavasta lähdetekstistä. Synonymian ja variaation selvittäminen vaatii sekin tekstikontekstia ja erikoisalakontekstia. Erikoisalan ulkopuolisen kontekstin tunteminen on tarpeen, kun on ratkaistava, onko yleiskielen sana tutkittavassa tekstissä termi ja ovatko termit monen alan yhteisiä termejä.

Kaikki kontekstitasot ovat siis välttämättömiä termintunnistuksessa, kun on päätettävä termiehdokkaiden termiydestä, vaikka termiehdokaslistan tuottaminen onkin mahdollista automaattisesti kvantitatiivisten ja lingvististen kriteerien perusteella. Käsitesuhteiden ja käsitepiirteiden poiminnassa tekstikonteksti on vieläkin tärkeämpää, sillä niiden tunnistamiseksi on tekstille tehtävä semanttinen analyysi. Yksittäisen tekstin käsitetieto ei kuitenkaan riitä kuvaamaan koko käsitettä, sillä teksteissä annetut määritelmät ovat ehkä vain yhden asiantuntijan näkemyksiä käsitteistä. Käsitteen kaikkien piirteiden selvittämiseksi tarvitaan monta tekstikontekstia, jolloin lähestytään käsitettä yksittäisen tekstikontekstin sijasta erikoisalakontekstin näkökulmasta. Yhdistämällä monen tekstin käsitetieto voidaan saada käsitteelle likimääräinen kuvaus, johon tosin vaikuttavat näkökulma ja mahdollisesti historialliset, sosiaaliset ja poliittiset seikat. Erikoisalan ulkopuolinen maailma onkin tekstissä aina mukana ja vaikuttaa tulkintaan (Croft & Cruse 2004: 103). Sosiaalisuus on mukana kielessä tai ainakin kielen merkityspotentialissa, joten kaikessa tutkimuksessa, myös terminologiassa, on välttämättä olemassa sosiaalinen aspekti (Pälli 2003: 56).

Olen edellä käsitellyt ensin termin kontekstista riippumattomia kvantitatiivisia ja lingvistisiä ominaisuuksia ja sitten kontekstin vaikutusta terminpoimintaan. Seuraavaksi selostan tutkimukseeni kuulunutta manuaalista ja puoliautomaattista terminpoimintaa.



## **6 MANUAALINEN JA PUOLIAUTOMAATTINEN TERMINPOIMINTA SUOMALAISISTA JA VENÄLÄISISTÄ MERENKULUN TURVALLISUUTTA KÄSITTELEVISTÄ TEKSTEISTÄ**

Johdannossa asetin tutkimukseni tavoitteeksi kehittää termien ja käsitetiedon poimintamenetelmiä aidoista käyttöteksteistä. Tavoitteen taustalla on oletus, että deskriptiivisessä sanastotyössä tarvittava terminologisesti kiinnostava tieto tai ainakin olennainen osa siitä on teksteissä. Tähän pääoletukseen liittyy oletus, että terminologisesti kiinnostava tieto on teksteissä usein ”piilossa” eli termit eivät esiinny teksteissä sanakirjamuodossa eikä käsitetietoa ole osoitettu eksplisiittisesti.

Tutkimukseni ensimmäinen osatavoite on terminpoimintamenetelmien kehittäminen. Termit voidaan poimia erikoisalan teksteistä periaatteessa kahdella tavalla. Käsitemenetelmässä voidaan edetä yläkäsitteistä alakäsitteisiin, jolloin lähdetään liikkeelle pienestä käsitejoukosta, jota laajennetaan, kunnes toivottu laajuus on saavutettu (induktiivinen menetelmä). Toinen mahdollisuus on aloittaa suuresta termiehdokasmäärästä, jolloin kohdealan teksteistä poimitaan kaikki mahdolliset termiehdokkaat määriteltäessä (deduktiivinen menetelmä). Käsiteanalyysiin ja sanastoon valitaan kuitenkin vain osa termiehdokkaista. (Nykänen 1999: 65; vrt. Martin 1992: 250.)

Tässä luvussa selostan terminpoimintaa, jossa termiehdokkaat poimittiin suomenkielisestä esitelmätekstistä ja venäjänkielisestä artikkelista kahdella deduktiivisella menetelmällä, manuaalisesti ja puoliautomaattisesti. Tavoitteena oli selvittää, miten erikoisalan opiskelijoista tai kieliaineiden opiskelijoista koostuvat koehenkilöryhmät ja tietokoneohjelmat selviävät terminpoiminnasta. Myös induktiivista menetelmää termiehdokaiden valinnassa voisi soveltaa manuaalisesti tai puoliautomaattisesti, mutta induktiivisen menetelmän testaus ja arviointi jää tämän tutkimuksen ulkopuolelle. Aloitan terminpoimintaosuuden käsittelyn kuvaamalla poiminnassa käytettyä aineistoa ja menetelmiä.

### **6.1 Terminpoiminnan aineisto ja menetelmät**

Tässä tutkimuksessa aineiston koostamiseen on kiinnitetty erityistä huomiota (ks. jakso 7.3). Pyrin valitsemaan monen hyvän tekstin joukosta terminpoimintaan parhaiten sopivat tekstit. Näin pyrin välttämään ongelmat, joihin Kageura, Yoshioko, Koyama ja Nozue (1998 lähteessä Soininen 1998: 18) viittaavat sanoessaan, että puoliautomaattisessa terminpöytäkirjassa ”termejä haetaan määrätyistä teksteistä, mutta tekstien alaan tai laatuun ei ole juuri kiinnitetty huomiota”. Ongelma liittyy erikoisalakorpusten kokoamisen työläyteen ja korpusten saatavuuteen. Kun valmiita korpuksia ei juuri ole ja korpusten kokoamiseen ei ole käytettävissä sen paremmin aikaa kuin työvoimaa, korpus koostetaan teksteistä, joita sattuu olemaan saatavilla. Korpuksen vaikutus tuloksiin on kuitenkin oleellisen tärkeä, sillä huonosta korpuksesta ei paraskaan ohjelma saa paljoa irti.

Terminpoiminnan aineistona käytin yhtä suomenkielistä esitelmää (pw02) ja yhtä venäjänkielistä artikkelia (Eskos). Termilistojen koostamiseen osallistuneet asiantuntijat ja koehenkilöinä toimineet opiskelijat poimivat termit samasta aineistosta. Valitsin tekstit niin, että niiden aihe, ikä, pituus ja erikoiskielisyys ovat mahdollisimman lähellä toisiaan. Tämä tavoite toteutuikin ilmeisen hyvin. Tekstien pituutta rajoittivat kokeessa käytetyt terminpoimintamenetelmät. Manuaalisessa poiminnassa tekstin on oltava riittävän lyhyt, jotta terminpoimijat eivät väsy kesken tehtävän. Tietokoneohjelmalla tehtävässä puoliautomaattisessa poiminnassa tekstin taas on oltava riittävän pitkä, jotta sanojen ja sanayhdistelmien frekvenssejä laskevat ohjelmat tuottavat vertailukelpoisen termiehdokaslistan. Pilottikokeiden perusteella noin 2 000 sanan teksti sopii sekä manuaa-



liseen että puoliautomaattiseen poimintaan eikä tekstin pituuden kasvattaminen paranna puoliautomaattisen poiminnan tuloksia. Seuraavassa taulukossa on esitetty terminpoiminnassa käytettyjen tekstien tärkeimmät piirteet terminpoimintatehtävän kannalta.

TAULUKKO 1. Terminpoiminnassa käytetyn aineiston kuvaus

	pw02	Eskos
Tekijä	Paavo Wihuri	О. Причкин & В. Ключев
Otsikko	Meriliikenteen ohjausjärjestelmät VTS ja VTMS	Единая система контроля над судоходством
Tekstilaji	Esitelmä	Artikkeli
Aihe	Meriliikenteen ohjausjärjestelmät	Meriliikenteen ohjausjärjestelmät
Julkaisuvuosi	2002	2000
Sivumäärä	19 (riviväli 2, kirjasinkoko 12)	22 (riviväli 2, kirjasinkoko 12)
Sanaluku	2 366	2 794

Terminpoiminnan tulosten vertailua varten tarvittavien termilistojen koostamiseen osallistui suomalaisia ja venäläisiä merenkulun asiantuntijoita ja terminologisia menetelmiä tuntevia henkilöitä. Suomenkielisestä tekstistä (pw02) termit poimi yksi merenkulun teksteihin erikoistunut kääntäjä ja kaksi terminologian tutkijaa. Venäjänkielisinä asiantuntijoina toimivat kaksi kansipäällystöön kuuluvaa merenkulkijaa ja kaksi terminologian tutkijaa, jotka poimivat termit venäjänkielisestä tekstistä (Eskos). Suomen- ja venäjänkieliset termilistat ovat terminpoimintakoettani varten luotuja normeja, joihin vertaan toisaalta käsinpoiminnan ja toisaalta puoliautomaattisen terminpoiminnan tuloksia. Normin luomiseen osallistuneista käytän tutkimuksessani nimitystä *asiantuntija*.

Koehenkilöinä oli suomalaisia ja venäläisiä merenkulun opiskelijoita sekä suomalaisia ja venäläisiä kieliaineiden opiskelijoita. Kukin koehenkilö poimi termiehdokkaita vain omaa äidinkieltään edustavasta tekstistä. Suomalaiset ja venäläiset kieliaineiden ja merenkulun opiskelijat muodostivat äidinkielen ja opiskeltavan alan mukaan jaettuna neljä ryhmää. Kustakin ryhmästä tehtävän palautti vähintään 11 koehenkilöä, joten otin vertailuun jokaisesta ryhmästä 11 terminpoimintaa. Niistä ryhmistä, joissa palautettuja tehtäviä oli tätä enemmän, valitsin vertailuun tulevat vastaukset satunnaisotannalla.

Suomalaisina koehenkilöinä toimivat ensinnäkin Helsingin yliopiston käännöstieteen laitoksen opiskelijat, jotka osallistuivat terminologian kurssiin syksyllä 2003. Heistä osa oli vasta aloittanut yliopisto-opinnot ja osa oli opiskellut jo muutaman vuoden. Käytän kääntämisen opiskelijoiden koehenkilöryhmästä myöhemmin lyhennettä *kofi*. Kokeeseen osallistuneet suomalaiset merenkulun opiskelijat olivat Kymenlaakson ammattikorkeakoulun opiskelijoita, jotka suorittivat merenkulun erikoistumisopinnot kuuluvan Merenkulun hallinto ja johtaminen -kurssin syksyllä 2003. Tästä koehenkilöryhmästä käytän lyhennettä *mofi*.

Venäläisinä koehenkilöinä toimivat Pietarin valtionyliopistossa syksyllä 2003 soveltavaa kielitiedettä opiskelleet, joiden muodostamasta koehenkilöryhmästä käytän myöhemmin lyhennettä *koru*. Venäläisten merenkulun opiskelijoiden koehenkilöryhmä koostui Makarovin meriakatemian kolmannen vuosikurssin kansipäällystöopiskelijoista. Tästä ryhmästä käytän lyhennettä *moru*. Myös venäläiset merenkulun opiskelijat osallistuivat terminpoimintakokeeseen syksyllä 2003.

Pyysin koehenkilöinä toimineita opiskelijoita poimimaan termit annetusta tekstistä (pw02 tai Eskos) alleviivaamalla sanat, sanaliitot tai sanayhdistelmät, jotka poimijan

mielestä ovat termejä. Koehenkilöille annettu ohje (liite 2 ja liite 3) oli hyvin yleisluontoinen, koska en halunnut johdatella terminpoimintaa liikaa. Pitkiä ohjeita ei muutoinkaan yleensä lueta. En pyytänyt koehenkilöitä jaottelemaan termejä varmoihin ja epävarmoihin tai minkään muun luokitteluperusteen mukaan vaan ainoastaan poimimaan termit alleviivaamalla. Monisanaisista termeistä pyysin alleviivaamaan toisella alleviivauksella ne sanat, jotka koehenkilön mielestä ovat eri termi kuin aiemmin alleviivattu samoja sanoja sisältävä pitempi tai lyhyempi termi.

Suomenkieliset merenkulun opiskelijat ja venäjänkieliset kieliaineiden opiskelijat tekivät terminpoimintatehtävän luokassa. Merenkulun opiskelijoilla tehtävään oli käytettävissä yksi oppitunti. Muut koehenkilöryhmät saivat tehdä tehtävän kotona ilman aikarajoitusta. Suomalaiset merenkulun opiskelijat olivat tehtävää tehdessään suorittamassa ammattiinsa liittyviä jatko-opintoja. He kaikki olivat jo työelämässä merellä tai maissa, merenkulku- tai satamahallinnossa tai varustamossa. Koehenkilöiltä ei kysytty ikää, mutta koska suomalaiset merenkulun opiskelijat olivat jo jatko-opintovaiheessa, he olivat myös iältään keskimäärin muiden opiskelijaryhmien koehenkilöitä vanhempia.

Opiskelijoiden palauttamista teksteistä kokosin ensin kaikki valitut termiehdokkaat Excel-taulukkoon. Palautin sanat perusmuotoon, ja poistin termiehdokkaaseen mahdollisesti liittyneet sulkeet tai lainausmerkit, koska alleviivausten perusteella ei aina voinut päätellä, oliko koehenkilö tarkoittanut niiden kuuluvan termiin vai ei. Sulkujen ja lainausmerkkien tehtävää terminologiassa käsitellään tässä tutkimuksessa terminologisten koettimien yhteydessä luvussa 7. Laskin eri termiehdokkaiksi sanaliitot, jotka oli valittu sekä erikseen että yhdessä, koska myös käytettävissä olleet termintunnistusohjelmat hakevat termiehdokkaita tällä tavalla. Tämän jälkeen aakkostin termiehdokkaat ja yhdistin monta kertaa valitut termiehdokkaat yhdeksi valinnaksi. Näin sain selville koehenkilöiden valitsemien eri termiehdokkaiden lukumäärän. Tämän jälkeen yhdistin opiskelijoiden termiehdokaslistat ryhmittäin siten, että vertailuun tulivat vähintään neljän samaan koehenkilöryhmään kuuluneen opiskelijan valitsemat termiehdokkaat, koska näin eri termien lukumäärä oli lähellä ryhmän keskiarvoa (vrt. Fulford 2001: 270).

Saadakseni selville, onko erikoisalaa opiskelevien ja kieliaineita opiskelevien terminvalinnassa eroja, jotka selittyvät erikoisalan tuntemuksella tai sen puutteella, vertasin kielittäin koehenkilöryhmien valintaa asiantuntijoiden valintaan. Samantapaisia psykologivistisiä kokeita on tehty ainakin Venäjällä. Esimerkiksi Gorodetskaja (1993: 31) vertasi, kuinka erikoisalan asiantuntijat ja maallikot ymmärtävät kahdesta sanasta muodostuvat sanaliittotermit.

Manuaalisen terminpoiminnan lisäksi termiehdokkaat tunnistettiin suomenkielisestä tekstistä Conexor Oy:n NaviTerm 2.0 -ohjelmalla ja Masterin terminpoimintaohjelmalla. Tradoksen MultiTerm Extract -ohjelmalla termit tunnistettiin sekä suomenkielisestä että venäjänkielisestä koetekstistä. Puoliautomaattisen termintunnistuksen tulosta vertasin kielittäin asiantuntijaryhmien poimintatulokseen (ks. jaksot 6.7.3 ja 6.7.4).

Tutkimuksessani koehenkilöryhmien ja automaattisten termintunnistusohjelmien tuottamia termiehdokaslistoja on arvioitu kriittisesti ja termeiksi on hyväksytty vain termiehdokkaat, jotka ovat morfologiselta ja syntaktiselta rakenteeltaan täysin samoja kuin asiantuntijoiden valitsemat termit. Esimerkiksi liian pitkiä termiehdokkaita ei ole hyväksytty tunnistettujen termien listaan, vaikka monet termintunnistusohjelmat tuottavat sanaliitoista eripituisia termiehdokkaita (vrt. L’Homme ym. 1996: 306). Varsin tiukkaa arviointikriteeriä perustelen sillä, että eripituiset termiehdokkaat saattavat hyvinkin olla termejä, mutta saattavat myös nimetä aivan eri käsitteitä.

Koehenkilöiden tekemän manuaalisen terminpoiminnan ja tietokoneohjelmien avulla tehdyn puoliautomaattisen terminpoiminnan onnistumista arvioin saannin ja tarkkuuden avulla (ks. jakso 6.2). Arviointia varten olen jakanut termiehdokkaat kolmeen ryh-

mään: termeihin, piilotermeihin ja valetermeihin. Jaottelu on yksinkertaistettu muoto Fulfordin (2001: 270) käyttämästä termiehdokkaiden jaottelusta. Fulford vertasi puoliautomaattisen terminpoiminnan tuottamia termiehdokaslistoja asiantuntijan käsin poimimien termien listaan. Vertailussaan hän jaotteli puoliautomaattisen terminpoiminnan tuottamat termiehdokkaat viiteen ryhmään. Yhden ryhmän muodostivat asiantuntijan poimimien termien kanssa täysin samat termiehdokkaat (term match), toisen ryhmän asiantuntijan poimimia termejä lyhyemmät (term truncation), kolmannen ryhmän niitä pidemmät termiehdokkaat (term expansion), neljännen ryhmän piilotermit (under-generation) ja viidennen ryhmän valetermit (overgeneration). Saannin ja tarkkuuden selvittämisen lisäksi tarkastelen eri menetelmillä poimittujen ja eri koehenkilöryhmien poimimien termien, valetermien ja piilotermin ominaisuuksista niiden pituutta, frekvenssiä ja termikaavoja. Terminpoimintatulosten arvioinnin yhteydessä selostan myös variaation esiintyvyyttä, terminvalinnan yksimielisyyttä ja terminpoimintaohjelmien antamien termiyspisteiden jakautumista.

Terminpoimintakoe perustui olettamukselle, että merenkulun tai terminologian asiantuntijat, opiskelijoista kootut koehenkilöryhmät ja termintunnistukseen tarkoitetut tietokoneohjelmat käyttävät valinnan tukena erilaista tietoa. Oletuksena oli, että merenkulun opiskelijoilla on merenkulun tuntemusta mutta ei terminologian teorian tuntemusta ja että kieliaineiden opiskelijoilla on terminologian teoriasta vähintään perustiedot mutta ei merenkulun tuntemusta. Lisäksi asiantuntijoilla ja koehenkilöillä on maailmantietoa eli asuinpaikasta, koulutuksesta ja ammatista riippumatonta ihmisille yhteistä tietoa maailmasta. Erikoisalalla työskentelevät ovat hankkineet koulutuksen ja työkokemuksen avulla erikoisalatietoa. Se on erikoisalalla toimivien yhteistä tietoa, jota muilla erikoisaloilla toimivilla ei ole. Tiettyä kieltä äidinkielenään puhuvalla on kielellistä tietoa eli tietoa kielen leksikaalisten yksiköiden ja syntaktisten rakenteiden muodosta, merkityksestä ja asianmukaisesta käytöstä. Terminologisia menetelmiä tuntevalla on kielestä lisäksi teoreettista tietoa. Käytettävissä oleva tieto voidaan esittää seuraavan luettelon muodossa (vrt. jakso 5.5):

#### **Erikoisalan asiantuntijan ja erikoisalaa opiskelevan käytössä oleva tieto**

- Maailmantieto
- Erikoisalatieto
- (Kielellinen tieto)
- Tekstikonteksti

#### **Kielen asiantuntijan ja kieliaineiden opiskelijan käytössä oleva tieto**

- Maailmantieto
- (Erikoisalatieto)
- Kielellinen ja terminologinen tieto
- Tekstikonteksti

#### **Tietokoneohjelman käytössä oleva tieto**

- Sanojen ja sanayhdistelmien frekvenssit
- Kieliopilliset rakenteet
- Sulkulista

Olettamuksen mukaan erikoisalaa tunteva tai opiskeleva perustaa valintansa maailmantietoon, erikoisalatietoon ja tietoon ympäröivästä tekstistä. Kielen asiantuntija tai opiskelija taas käyttää maailmantietoa, kielellistä ja terminologista tietoa sekä tietoa ympäröivästä tekstistä. Erikoisalatietoa ja kielellistä tietoa ei kuitenkaan voi pitää pelkästään

tietyt koehenkilöryhmän käytössä olevana keinona, koska henkilöillä, jotka eivät toimi tietyllä erikoisalalla, on erikoisalasta maallikon tiedot, jotka on saatu esimerkiksi tiedotusvälineitä seuraamalla, ja erikoisalan asiantuntijat ovat todennäköisesti hyvinkin tottuneita kielenkäyttäjiä. Kielen asiantuntijoilla ja kieliaineita opiskelevilla on tarkasteltavasta erikoisalasta todennäköisesti vain maallikon tiedot, mutta heillä on terminologian erikoisalatietoa, jota he käyttävät terminpoiminnassa. Ero erikoisalan asiantuntijoiden ja terminologisia menetelmiä tuntevien välillä on juuri tämä: asiantuntija poimii termejä, joiden nimeämät käsitteet ovat hänelle tuttuja erikoisalan kokemuksen perusteella, ja terminologi kerää termejä perustaan valintansa siihen, millaisia termit yleensä ovat.

Terminpoiminnan koeasetelma perustuu oletukselle, että erikoisalan opiskelijoiden ja kieliaineiden opiskelijoiden termiehdokaslistoissa on eroja, jotka selittyvät toisaalta erikoisalan ja toisaalta terminologian teorian tietämyksellä tai sen puutteella. Koehenkilöinä toimineet kieliaineiden opiskelijat eivät oletettavasti tunnista kaikkia erikoisalan termejä ja valitsevat termeiksi sanoja ja sanayhdistelmiä, jotka eivät ole termejä tai ovat jonkin toisen erikoisalan termejä. Merenkulun opiskelijat puolestaan poimivat oletettavasti termiehdokkaita, jotka eivät voi olla termejä kielellisen rakenteensa vuoksi.

Tietokoneohjelma laskee yksittäisten sanojen ja sanayhdistelmien esiintymät ja mahdollisesti tunnistaa kieliopillisen rakenteen (esim. adj. + subst.). Terminpoimintaohjelmassa voi olla myös sulkulista, johon on koottu sellaiset merkkijonot, jotka eivät voi olla termejä tai termin osia.

## 6.2 Terminpoimintatulosten vertailussa käytetyt mittarit

Terminpoimintamenetelmien arviointia vaikeuttavat monet terminologian teorian ongelmat. Kageura ym. (1998: 81–85) sanovat, että arviointia haittaa ensinnäkin poiminnan kohteen eli termin epämääräisyys. Tämä piirre liittyy termin käsitteen ongelmallisuuteen, jota olen käsitellyt jaksossa 2.3. Kielen monimuotoisuudesta ja dynaamisuudesta johtuen käsite *termi* todennäköisesti jääkin rajoiltaan hämäräksi. Toiseksi terminpoiminnan tulokseen vaikuttavat tietyt lähtöaineiston piirteet. Tässä tutkimuksessa on eri menetelmin toteutetuissa terminpoiminnoissa käytetty samaa lähtöaineistoa, jolloin tulokset ovat tältä osin vertailukelpoisia. Kolmanneksi terminpoiminnan tulosten arviointiin ei ole yleisesti käytettyjä arviointikriteerejä, vaan vertailukohtana on yleensä manuaalisen terminpoiminnan tuottama termilista, joka on aina subjektiivinen, vaikka poimija olisi erikoisalan asiantuntija tai terminologi. Tutkimuksessani subjektiivisuutta on pyritty vähentämään koostamalla vertailulista merenkulun asiantuntijoiden ja terminologien yhteistyönä. Mainitut teoreettiset ongelmat eivät hankaloita pelkästään puolipäätöksen, vaan yhtä lailla myös manuaalisen terminpoiminnan arviointia sekä terminpoimintamenetelmien vertailtavuutta.

Terminpoiminnan tulosten arviointiin vaikuttaa myös poiminnalle asetettu tavoite. Terminologin tavoitteena saattaa olla täydellinen terminpoiminta, jolloin tulokseen halutaan kaikki termit. Kääntäjä hakee ehkä vain yhdelle käsitteelle vastinetta ja kontekstia. Arviointia ei saisikaan perustaa vain määrällisille kriteereille, vaan poimijan tyytyväisyys tulokseen pitäisi asettaa yhdeksi arviointiperusteeksi (Blair 1990: 9–13). Tyytyväisyys on kuitenkin suhteellista: poimija voi olla tyytyväinen määrällisillä mittareilla mitattuna huonoon tulokseen, koska ei tiedä, mitä on jäänyt paitsi. Toisaalta poimija voi olla tyytymätön tulokseen, jos lähtötekstissä ei ollut toivottua tietoa, vaikka määrällisesti arvioituna poiminta voi olla hyvinkin onnistunut. Kiinnostavuudestaan huolimatta käyttäjätyytyväisyyden mittaaminen on rajattu tämän tutkimuksen ulkopuolelle.

Terminpoiminnan määrällisten tavoitteiden arvioinnissa on terminologisissa tutkimuksissa yleisesti käytetty tiedonhausta (informational retrieval) lainattuja mittareita

*tarkkuus* (precision) ja *saanti* (recall) (ks. esim. Strehlow 2001: 428). Tarkkuus on poimittujen termien määrä jaettuna poimittujen termiehdokkaiden määrällä, ja saanti on poimittujen termien määrä jaettuna kaikkien tekstissä olevien termien määrällä. Tarkkuus lasketaan kaavasta  $P = a/(a+b)$  ja saanti kaavasta  $R = a/(a+d)$ , joissa  $a$  = poimittujen termien lukumäärä,  $b$  = poimittujen ei-termien lukumäärä ja  $d$  = poimimatta jääneiden termien lukumäärä. Jos luvut halutaan ilmoittaa prosentteina, on edellä esitetyt suhdeluvut kerrottava sadalla. Tarkkuuden ja saannin mittaaminen edellyttää siis, että poimintatuloksen arvioijalla on käytössään vertailulista, johon poimintatulosta verrataan.

Koska saanti ja tarkkuus ovat vastakkaisia, maksimaalista saantia ja maksimaalista tarkkuutta ei voida saavuttaa samanaikaisesti, vaan on löydettävä määrän ja laadun tasapaino sen mukaan, kumpaa tavoitetta pidetään tärkeämpänä. Jos pyritään mahdollisimman suureen saantiin eli tavoitteena on hakea kaikki termit, terminhaun tarkkuus kärsii eli haku tuottaa runsaasti osumia, jotka eivät ole termejä. Jos taas pyritään välttämään turhia hakutuloksia, tarkkuus kärsii, sillä osa termeistä jää hakutuloksen ulkopuolelle. (Ks. esim. Lahtinen 2000: 62.)

Tarkkuuden ja saannin lisäksi on varsinkin puoliautomaattisen terminpoiminnan arvioinnissa käytetty myös muita mittareita, kuten esimerkiksi haun kestoa (Strehlow 2001: 428). Tämän tutkimuksen terminpoimintakokeessa ei kuitenkaan mitattu terminpoimintatehtävään käytettyä aikaa, joten en ole käyttänyt haun kestoa arviointikriteerinä.

Menetelmän hyvyttä ja huonoutta voidaan arvioida sen käytettävyydellä, joka voidaan määritellä helppokäyttöisyyden, tehokkuuden ja miellyttävyyden summaksi (Parkinen 2002). Ratkaisevaa ei ole ainoastaan se, pystytäänkö menetelmällä tekemään halutut toimenpiteet, vaan myös se, miten helppo tai miellyttävä menetelmä on käyttää. Esimerkiksi tietokoneohjelman valintaan vaikuttavat käyttöliittymän käyttäjäystävällisyys, ohjelman yhteensopivuus yrityksen tai käyttäjän työympäristöön ja ohjelman hinta (Jaakkola 2002: 72). Nämä kaikki tekijät vaikuttavat ratkaisevasti siihen, mikä vaihtoehtoisista ohjelmista valitaan tai valitaanko ohjelmaa ollenkaan. Tässä tutkimuksessa keskityn tietokoneohjelmien arvioinnissa kuitenkin vain ohjelmien tuottamien termiehdokaslistojen arviointiin ja muita ohjelmien käyttökelpoisuuteen vaikuttavia ominaisuuksia sivuan vain pikaisesti.

Terminpoiminnan tulosta voidaan arvioida myös kvalitatiivisesti, jolloin mittareina käytetään pysyvyyttä (unithood) ja termiyyttä (termhood). Kageuran ja Uminon (1996: 260–261) mukaan pysyvyys viittaa sanaliittojen tai kollokaattien yhteenkuulumisen voimakkuuden tai vakauden asteeseen. Termiys puolestaan viittaa asteeseen, jolla kielellinen yksikkö liittyy erikoisalan käsitteeseen. Pysyvyyttä voidaan mitata variaation yleisyydellä eli termiehdokkaan varianttien lukumäärällä. Termiyyttä voidaan mitata valinnan yksimielisyydellä manuaalisessa terminpoiminnassa ja puoliautomaattisten terminpoimintaohjelmien termiehdokkaille antamalla termiyspisteillä. Olen käyttänyt variaation yleisyyttä, terminvalinnan yksimielisyyttä ja termiyspisteitä oman tutkimukseni terminpoiminnan arvioinnissa, jonka tulokset olen koonnut jaksoon 6.8.

### 6.3 Suomalaisen ja venäläisen termilistan laadinta

Tutkimukseni terminpoimintakokeen tarkoituksena oli verrata koehenkilöryhmien ja terminpoimintaohjelmien tuottamia termiehdokaslistoja normiin eli suomalaisten ja venäläisten termien listoihin. Niiden tuottaminen uskottiin merenkulun tai terminologian asiantuntijoille. Yhtä useamman asiantuntijan käytöllä pyrin vähentämään valinnan subjektiivisuutta, sillä kokemuksen mukaan eri henkilöiden samasta tekstistä tuottamat termilistat poikkeavat toisistaan varmasti (ks. esim. Laakso 2003: 27). Tämän ilmiön

vahvasti todeksi ennen varsinaista terminpoimintakoetta tekemäni pilottikoe, jonka tavoitteena oli tutkia yksilöllisiä eroja terminvalinnassa. Kokeeseen osallistui kolme suomalaista merikapteenin koulutuksen saanutta merenkulun asiantuntijaa.

Pilottikokeeseen osallistuneiden merenkulun asiantuntijoiden terminvalinnassa oli huomattavia yksilöiden välisiä eroja koulutuksen, sukupuolen, iän ja työkokemuksen samankaltaisuudesta huolimatta. Seuraavaan taulukkoon on koottu merikapteenien valitsemien ja oman valintani mukaisten eri termiehdokkaiden määrät ja pituudet sanaluvulla mitattuna.

TAULUKKO 2. Pilottikokeeseen osallistuneiden merikapteenien termiehdokkaiden poiminta Meriteiden säännöistä

Sanaluku	mk1		mk2		mk3		Oma valinta	
	f	%	f	%	f	%	f	%
1	44	34,7	14	10,2	76	21,5	175	51,8
2	47	37,0	19	13,8	88	24,9	99	29,3
3	24	18,9	22	15,9	86	24,4	28	8,3
4	7	5,5	18	13,0	42	11,9	13	3,8
≥ 5	5	3,9	65	47,1	61	17,3	23	6,8
Yhteensä	127	100	138	100	353	100	338	100

Koehenkilö mk1 valitsi sanoja ja kahden tai kolmen sanan sanaliittoja, jotka hyvin todennäköisesti myös ovat termejä. Tämän koehenkilön valitsemien eri termiehdokkaiden lukumäärä on kuitenkin vain noin kolmannes omasta valinnastani, joten todennäköisesti koehenkilö mk1 ei valinnut kaikkia termejä. Koehenkilö mk2 valitsi lähes saman määrän eri termiehdokkaita kuin koehenkilö mk1, mutta hänen valitsemansa termiehdokkaat, esimerkiksi *ei saa estää*, ovat pikemmin merenkulun fraaseja kuin termejä. Koehenkilö mk2 valitsi hyvin usein myös kokonaisia lauseita, esimerkiksi *jokaisen aluksen on kuljettava turvallisella nopeudella*. Koehenkilön mk3 valitsemien termiehdokkaiden lukumäärä on hyvin lähellä itse valitsemieni termien lukumäärää, mutta suuri osa termiehdokkaista on huomattavan pitkiä sanaluvulla mitattuna. Samoin kuin koehenkilö mk2 myös koehenkilö mk3 valitsi termeiksi fraaseja, esimerkiksi *auringonnoususta auringonlaskuun*. Huomattava osa näiden kahden koehenkilön termiehdokkaista onkin yli viiden sanan mittaisia. Suuri osa kahden koehenkilön termiehdokkaista ei siis noudattanut termien kielellisiä vaatimuksia.

Mahdollinen selitys pilottikokeen terminvalinnan eroihin on epäonnistunut koetektin valinta. Pearsonin (1998: 27) mukaan termeiksi valitaan nimittäin yleensä oudot sanat, joiden merkitystä ei tiedetä, sekä teknisiltä näyttävät sanat. Pilottikokeessa aineistona oli noin 5 000 sanan ote säädöstekstistä *Kansainväliset säännöt yhteentörmäämisen ehkäisemiseksi merellä, 1972* (SopS n:o 30/1977), jonka suomalaiset merenkulkijat tuntevat nimellä *Meriteiden säännöt*. Se on eräs tärkeimmistä kansipäällystön koulutuksessa opiskeltavista säädöksistä. Tekstissä tuskin oli koehenkilöille outoja termejä, mutta sitäkin enemmän lauseita, jotka opintojen aikana on opiskeltu ulkoa. Koska yksi koehenkilö kuitenkin valitsi termiehdokkaita huomattavasti muista poikkeavasti, ei pelkkä koulutus selitä koehenkilöiden valintojen välistä eroa. Ehkäpä lyhyitä termiehdokkaita valinnut koehenkilö käytti avustajaa.

Erikoisalan asiantuntijan tekemää terminpoimintaa pidetään yleensä luotettavampana kuin terminologin tekemää terminpoimintaa (ks. esim. Fulford 2001). Pilottikoe ei kuitenkaan antanut kovin rohkaisevia tuloksia erikoisalan asiantuntijoiden käytöstä ter-



minhaussa, jos asiantuntijoilla ei ole minkäänlaisia perustietoja terminologiasta. Lyhyt koulutus ennen tehtävän suorittamista olisi voinut parantaa tulosta, sillä tuloksia tarkemmin tutkimalla paljastuu, että asiantuntijoiden valinnat ovat hyvin samankaltaiset, kun koehenkilöiden alleviivaamista ilmauksista jätetään ”ylimääräiset” sanat pois. Seuraavaan taulukkoon on koottu esimerkkejä erimittaisista termiehdokkaista, joissa ytimeenä on sama termi.

TAULUKKO 3. Esimerkkejä pilottikokeen koehenkilöiden poimimista erimittaisista termiehdokkaista

mk1	mk2	mk3
avomerellä	koskevat kaikkia aluksia avomerellä	aluksia avomerellä
yhteentörmäämisen vaara	ratkaistakseen, onko yhteentörmäämisen vaara olemassa	yhteentörmäämisen vaara olemassa
tutkahavainto	eikä varsinkaan puutteellisiin tutkahavaintoihin	puutteellisiin tutkahavaintoihin
hyvää merimiestapaa	hyvissä ajoin ja hyvää merimiestapaa noudattaen	hyvää merimiestapaa noudattaen
turvallinen välimatka	turvallisen välimatkan päässä	alus sivuutetaan turvallisen välimatkan päässä
liikennejakolinjasta	pysyttävä erillään liikennejakolinjasta	pysyttävä erillään liikennejakolinjasta ja -vyöhykkeestä

Taulukon 3 esimerkit osoittavat, että koehenkilöt ovat paikallistaneet tekstistä termit, mutta eivät ole yksimielisiä siitä, mitkä sanat niihin kuuluvat. Jos asiantuntijoille olisi ennen koetta kerrottu, millaisia termit yleensä ovat muodoltaan, olisi kokeen tulos todennäköisesti ollut toisenlainen. Terminpoimijoille annettujen ohjeiden merkitys on havaittu muissakin tutkimuksissa (ks. esim. Kudashev 2007: 179).

Jotkut muutkin tutkijat ovat havainneet, että asiantuntijoiden käyttö terminpoiminnassa ei ole ongelmatonta. Karihalmeen (1996: 78) sanoin uudella alalla asiantuntijoiden käyttöä haittaavat ”tieteellinen hapuilu, erimielisyys ja tieteenulkoiset intressit”. Vivaldi ja Rodríguez (2001: 44) sanovat, että erikoisalan asiantuntijoilla on taipumus luokitella termeiksi vain tekstin aiheeseen selvästi liittyvät ilmaukset. Esimerkiksi *anemia* ei lääketieteen asiantuntijoiden mielestä ollut termi astmaa käsitelleessä tekstissä.

Vaikka ihmisten tekemä terminvalinta on parhaassakin tapauksessa vain hyvä kompromissi eikä niin sanottua oikeaa terminlistaa ehkä ole olemassakaan, päätin tutkimukseeni kuuluvassa terminpoimintakokeessa kuitenkin pyrkiä mahdollisimman lähelle ihannetta, jotta koehenkilöryhmien valinnan ja tietokoneohjelmien tuottamien termiehdokaslistojen vertailu asiantuntijoiden tuottamaan vertailulistaan tuottaisi perusteltuja tuloksia sattumanvaraisten tulosten sijasta.

Suomalaiset asiantuntijat kokosivat terminlistansa kolmessa vaiheessa ja venäläiset kahdessa vaiheessa tavoitettavuusongelmien vuoksi. Ensimmäisessä vaiheessa asiantuntijat poimivat termit koeteksteistä (pw02 tai Eskos) alleviivaamalla. Toisessa vaiheessa kokosin poimintatulokset kahdeksi termiehdokaslistaksi, joissa olivat kaikki asiantuntijoiden valitsemat termiehdokkaat. Koska tavoitteena oli koostaa sellaiset terminlistat, jotka kaikki asiantuntijat voisivat hyväksyä, pyysin asiantuntijoita käymään yhteisen termiehdokaslistan läpi ja tarkistamaan omaa valintaansa. Venäläisten asiantuntijoiden

termilistaan jäivät ne termiehdokkaat, joiden termiydestä kaikki neljä asiantuntijaa olivat yksimielisiä toisen vaiheen jälkeen.<sup>1</sup> Suomalaisille asiantuntijoille annoin toisen vaiheen jälkeen vielä listan niistä termiehdokkaista, joiden termiydestä he eivät olleet yksimielisiä. Pyysin heitä vielä kerran tarkistamaan valintaansa ja joko hyväksymään tai hylkäämään kiistanalaiset termiehdokkaat. Näiden kolmen vaiheen jälkeen suomalaiset asiantuntijat olivat koostaneet termilistan, jonka jokainen heistä voi hyväksyä.

Ensimmäisen vaiheen jälkeen suomalaisten asiantuntijoiden valitsemia termiehdokkaita oli 314. Kaikkien yksimielisesti valitsemia termejä on listassa 118 eli lähes 38 prosenttia termiehdokkaista. Tulosta voidaan pitää yllättävän hyvänä. Silti termiehdokkaiden joukossa oli ensimmäisen vaiheen jälkeen vielä 196 sellaista termiehdokasta, jotka vain yksi tai kaksi suomalaista asiantuntijaa oli valinnut. Näiden termiehdokkaiden osuus kaikista termiehdokkaista oli 62,4 prosenttia. Toisen vaiheen jälkeen kaikkien kolmen suomalaisen asiantuntijan valitsemien termien määrä oli noussut 172:een. Jäljellä oli vielä 32 ensimmäisessä vaiheessa valittua termiehdokasta, joiden termiys ei ollut varmistunut toisessa valintavaiheessa. Kolmannessa vaiheessa näistä 17 hyväksyttiin termeiksi, ja neuvottelun tuloksena oli yhteinen termilista, jossa on 189 termiä (Liite 4).

Venäläiset asiantuntijat valitsivat koetekstistä *Eskos* kaikkiaan 317 eri termiehdokasta. Yksimielisesti valittuja oli kuitenkin vain 37 eli 11,7 prosenttia. Venäläisten asiantuntijoiden väliset yksilölliset erot olivat siis huomattavasti suuremmat kuin suomalaisten asiantuntijoiden. Eniten termiehdokkaita valinnut asiantuntija valitsi lähes kaksinkertaisen määrän termiehdokkaita verrattuna vähiten valinneen valintaan, ja kolmas asiantuntija sijoittuu termiehdokkaiden lukumäärien vertailussa lähelle kahden muun keskiarvoa. Toisen vaiheen jälkeen venäläisten asiantuntijoiden termilistassa oli 107 sellaista termiä, jotka kaikki kolme asiantuntijaa valitsi yksimielisesti (Liite 5).

Näin suuret erot jopa terminologian asiantuntijoiden terminvalinnassa tukevat yleistä käsitystä, että kielellisen aineksen jako termeihin ja ei-termeihin ei ole ongelmaton tehtävä. Kaikki asiantuntijoiden valitsemat termiehdokkaat olivat kuitenkin täysin mahdollisia termejä, jos niitä arvioidaan yleensä termintunnistuksessa käytetyillä kriteereillä, kuten esimerkiksi morfologisella ja syntaktisella rakenteella sekä kuulumisella erikoisalalle.

Joidenkin termiehdokkaiden termiydestä eivät asiantuntijat esittäneet yksiselitteisen varmoja mielipiteitä, joten muutaman termiehdokkaan suhteen jouduin tulkitsemaan vastaajien näkemyksiä. Epävarmoissa tapauksissa termiehdokas on otettu listaan. Vastaajat myös saattoivat muuttaa mielipidettään kokeen aikana. Eniten epävarmuutta oli suomalaisten asiantuntijoiden kesken EU:hun ja koulutukseen ja venäläisten asiantuntijoiden kesken tietotekniikkaan liittyvien termiehdokkaiden valinnassa.

Suomalaisten asiantuntijoiden alkuperäisestä termiehdokaslistasta karsiutui 125 termiehdokasta. Näitä termiehdokkaita kaikki kolme asiantuntijaa eivät olleet valmiita hyväksymään termeiksi. Karsiutuneet termiehdokkaat ovat projektien, työryhmien tai asiakirjojen nimiä (*COST 301-tutkimus*, *DG VII*) ja monen alan yhteisiä termejä (*kan-sainvälinen säännöstö*, *jatkokoulutus*).

Venäläisten asiantuntijoiden termiehdokaslistasta karsiutui 210 termiehdokasta. Karsiutuneet ovat yksilökäsitteiden nimiä (*Тихоокеанский бассейн*, *Токийский меморандум*) ja niistä muodostettuja lyhenteitä (*ТОИТЦ*, *ФПС*), tietotekniikan termejä (*банк данных*, *сервер*, *информационная технология*), yleiskielisiksi tulkittavia sanoja (*акватория*, *движение*, *залив*, *контроль*, *побережье*, *подсистема*, *подцентр*, *сер-*

---

<sup>1</sup> Yhden ensimmäisessä vaiheessa mukana olleen asiantuntijan arvioinnista vastasi venäjänkielinen asiantuntija, joka on saanut samanlaisen koulutuksen ja toimii samankaltaisissa työtehtävissä kuin alkuperäisen valinnan tehnyt asiantuntija.

вис, средство, уровень) tai valettermejä, jotka ovat lähinnä puhekielisiä termin muunnoksia (зона контроля, контроль за судами).

### 6.3.1 Suomalaiden asiantuntijoiden valitsemien termien ominaisuuksia

Tässä jaksossa tarkastelen suomalaisten asiantuntijoiden valitsemien termien piirteitä ja etsin vastausta kysymykseen, onko termeillä, joiden valinnassa asiantuntijat olivat yksimielisiä, sellaisia määrällisiä tai kielellisiä ominaisuuksia, jotka korostavat niiden termiä. Tarkastelen ensin termien pituutta sanoina ja termien frekvenssiä. Tekstistä pw02 on syytä mainita, että se on julkaisematon eikä sen kirjoitusasu ole tästä syystä täysin viimeistelty. Muun muassa osa englanninkielisen lyhenteen ja suomenkielisen substantiivin yhdistelmistä on kirjoitettu yhteen yhdysmerkin kanssa ja osa erikseen ilman yhdysmerkkiä. Tämä kirjoitustapa vaikuttaa sanaluvun lisäksi termikaavoihin, joita tarkastelen sanaluvun ja frekvenssin jälkeen. Jakson lopussa tarkastelen vielä lyhyesti suomalaisten asiantuntijoiden valitsemien termien mahdollista vierasperäisyyttä ja variaatiota.

Suomalaisten asiantuntijoiden valitsema termit ovat keskimäärin 1,54 sanan mittaisia, kun sanaksi lasketaan merkkijono, jonka kummallakin puolella on välilyönti. Lyhennetermit ja termit, joissa osien välissä on yhdysmerkki, on laskettu yksisanaiksi. Yli puolet (61,4 %) termeistä on yksisanaista, kaksisanaista on noin kolmannes (28,6 %) ja vähintään kolmisanaista vain 10 prosenttia. Ensimmäisellä kierroksella yksimielisesti valittuihin termiin verrattuna yksisanaisten termien osuus on laskenut, kaksisanaisten osuus pysynyt lähes ennallaan ja vähintään kolmisanaisten osuus kasvanut (ks. Taulukko 4).

TAULUKKO 4. Suomalaiden asiantuntijoiden poimimien termien pituus sanoina

Sanaluku	1. valintakierros (A=1,39)		Termilista (A=1,54)	
	f	%	f	%
1	80	67,8	116	61,4
2	33	28,0	54	28,6
3	2	1,7	12	6,3
≥ 4	3	2,5	7	3,7
Yhteensä	118	100	189	100

Eripituisten termien osuudet eivät siis juuri muuttuneet kahden neuvottelukierroksen jälkeen. Termilistaan tuli tasaisesti lisää yhden, kahden ja kolmen sanan mittaisia termejä, vaikka painopiste siirtyikin hieman yhden sanan termeistä kolmen ja neljän sanan termiin. Valmiissa termilistassa monisanaisten termien osuutta kasvattavat englanninkieliset termit, joita kolmisanaista termeistä on puolet eli kuusi, esimerkiksi termi *Geographical Information System*, ja vähintään nelisanaista samoin noin puolet eli kolme, esimerkiksi termi *Vessel Traffic Management and Information System*.

Asiantuntijoidenkaan ei aina ole helppo ratkaista, mistä termi alkaa ja mihin se päättyy (vrt. Ahmad & Rogers 2001: 754–755; Sewangi 2001: 11). Termin alussa saattaa olla joukko määritteitä, joista osa voi kuulua termiin ja osa ei. Ensimmäisellä kierroksella valituissa suomalaisissa termiehdokkaissa vaihtelua oli esimerkiksi toista substantiivisia määrittävien substantiivien *alus* tai *laiva* valinnassa. Termilistaan olivat tarjolla esimerkiksi termiehdokkaat *laivan sijainti* ja *sijainti* sekä termiehdokkaat *aluksen tur-*

*vallisuus* ja *turvallisuus*. Neuvottelun jälkeen termeiksi jäivät *sijainti* ja *turvallisuus* ilman määritettä. Adjektiivimääritteissä vaihtelua oli paikkaa ilmaisevien adjektiivien *alueellinen*, *paikallinen*, *kansallinen* ja *kansainvälinen* kuulumisessa termiin. Ehdolla olivat esimerkiksi termiehdokkaat *alueellinen liikenne*, *paikallinen satama*, *kansallinen karttakorjauspalvelu*, *kansainvälinen säännös* ja *kansainvälinen säännöstö*. Termilistaan jäivät neuvottelun jälkeen termit *liikenne*, *satama* ja *karttakorjauspalvelu* ilman adjektiivimääritettä. Säännöstöä ja säännöstä ei termilistaan lopulta valittu sen paremmin määriteosan kanssa kuin sitä ilmankaan. Samansuuntaiseen tulokseen on päätynyt myös Soininen (1999a: 13–15). Hänen tekemässään sanastuskokeessa terminologien valitsemien termiehdokkaiden pituudessa oli vaihtelua, koska poimijat eivät olleet yksimielisiä adjektiivimääritteiden ja prepositioliusekkeiden kuulumisesta termiin.

Termin alussa tai lopussa olevien määritteiden lisäksi termin pituuteen vaikuttaa termien sisäkkäisyys. Esimerkiksi sanaliitto *alueellinen liikenteen informaatiopalvelu* sisältää kolme asiantuntijoiden valitsemaa termiä: *alueellinen liikenteen informaatiopalvelu*, *liikenteen informaatiopalvelu* ja *informaatiopalvelu*. Mahdollinen termi olisi myös sanaliitto *alueellinen informaatiopalvelu*, mutta sitä eivät asiantuntijat ole valinneet termilistaan.

Suomalaisten asiantuntijoiden valitsemat termit esiintyvät lähtötekstissä keskimäärin 2,34 kertaa. Vain kerran tekstissä esiintyviä termejä on 65,6 prosenttia. Luvussa on mukana myös 15 sellaista termiä, jotka eivät esiinny tekstissä kertaakaan siinä muodossa, jossa ne ovat termilistassa. Kahden ja vähintään kolmen esiintymän termejä on kumpiakin alle viidennes. Ensimmäisellä kierroksella yksimielisesti valittuihin termeihin verrattuna varsinkin yhden esiintymän termien osuus on kasvanut eli neuvottelukierrosten jälkeen asiantuntijat ovat lisänneet termilistaan erityisen monta harvoin esiintyvää termiä (ks. Taulukko 5).

TAULUKKO 5. Suomalaisten asiantuntijoiden poimimien termien frekvenssi

Frekvenssi	1. valintakierros (A=2,98)		Termilista (A=2,34)	
	f	%	f	%
≤ 1	71	60,2	124	65,6
2	20	16,9	35	18,5
3–4	8	6,8	10	5,3
≥ 5	19	16,1	20	10,6
Yhteensä	118	100	189	100

Vain kerran esiintyvien termien osuus on lopullisessa listassa hieman suurempi kuin heti yksimielisesti valittujen termien listassa, mutta harvoin esiintyvien termien osuus on kummassakin listassa reilusti yli puolet. Korkea frekvenssi ei siis ole termin välttämätön ominaisuus.

Terminpoiminnan näkökulmasta on hyödyllistä tarkastella termin pituuden ja frekvenssin lisäksi termin osien sanaluokkia, koska moni puoliautomaattinen terminpoimintaohjelma tunnistaa termit niiden osien sanaluokkien perusteella. Käytän tässä tutkimuksessa termien osien sanaluokkarakenteesta nimitystä *termikaava*. Termilistassa on runsaasti englanninkielisiä termejä ja lyhennetermejä, jotka on muodostettu englanninkielisistä sanoista. Tarkastelen lyhenteitä erikseen, koska niitä käytetään merenkulun suomessa ikään kuin suomen sanoina. Niihin liitetään suomen kielen mukaisia sijapäätteitä, ja ne muodostavat sanoja tai sanaliittoja suomenkielisten sanojen kanssa. Englanninkielisistä termeistä en ole muodostanut termikaavoja, koska en tutki niitä.

Tutkimukseni terminpoimintakokeeseen osallistuneet suomalaiset asiantuntijat olivat yksimielisiä termien ja terminosien sanaluokasta: kaikki ensimmäisessä vaiheessa valitut termit ovat substantiiveja tai substantiivilausekkeita. Ensimmäisellä kierroksella valituista termeistä noin kolmannes on yhdyssubstantiiveja eli muodostettu termikaavalla  $N_c$ , esimerkiksi *laivaliikenne*. Seuraavaksi eniten on lyhenteen ja substantiivin muodostamia termejä Abb-N, jossa lyhenteen ja substantiivin välillä on yhdysmerkki, esimerkiksi *VTS-operaattori*, tai Abb N, jossa lyhenteen ja substantiivin välillä ei ole yhdysmerkkiä, esimerkiksi *VTS operaattori*. Tämä suomen kielen oikeinkirjoitussääntöjen vastainen kirjoitustapa lienee levinnyt englannin kielen vaikutuksesta. Olen luokitellut yhdysmerkin kanssa kirjoitetut lyhenteen ja substantiivin yhdistelmät yhdyssanoiksi ja ilman yhdysmerkkiä kirjoitetut kaksisanaisiksi sanaliitoiksi. Yhdyssubstantiivien ja lyhenteen ja substantiivin yhdistelmien jälkeen lukumäärältään suurin ryhmä ovat muut substantiivit kuin yhdyssubstantiivit, esimerkiksi *luotsi*. Niiden termikaava on N. Kahden substantiivin tai adjektiivin ja substantiivin muodostamia sanaliittotermejä  $N_{(c)}$   $N_{(c)}$  tai A  $N_{(c)}$ , esimerkiksi *laivaliikenteen ohjaus* tai *elektroninen tietojenvaihto*, on kumpiakin noin 10 prosenttia. Suluissa oleva pieni c-kirjain termikaavassa tarkoittaa, että substantiivi voi olla yhdyssubstantiivi, kuten *laivaliikenne* ja *tietojenvaihto* edellä annetuissa esimerkeissä. Lisäksi listassa on kuusi lyhennetermiä, esimerkiksi *VTS*. Tässä yhteydessä on syytä huomauttaa, että variaation vuoksi yhdellä käsitteellä voi tekstissä olla useampia nimityksiä kuin yksi, joista jokin voi olla esimerkiksi yhdyssana tai kahden substantiivin sanaliitto. Näin samaa käsitettä nimeävät termit voivat noudattaa erilaisia termikaavoja, esimerkiksi  $N_c$  ja N N.

Toinen ja kolmas vaihe eivät juuri tuoneet muutosta sanaluokkakajakaumaan, sillä lopullisessa suomalaisten asiantuntijoiden termilistassa on vain substantiiveja, substantiivilausekkeita ja yksi verbitermi, *ohjailla*. Samansuuntaisia tuloksia on saanut esimerkiksi Soininen (1999a: 13–15), jonka terminpoimintakokeeseen osallistuneet terminologit poimivat vain substantiivilausekkeita. Lopulliseen listaan valitut uudet termit tuottivat viisi uutta termikaavaa, mutta niiden mukaisia termejä on vain kuusi. Lopullisen listan termeistä peräti 96,8 prosenttia noudattaa ensimmäisellä kierroksella valittujen termien kahdeksaa termikaavaa. Tavallisin termin rakenne on edelleen yhdyssubstantiivi. Lyhenteen ja substantiivin muodostamien termien lukumäärä pysyi samana, jolloin niiden suhteellinen osuus laski. Muuten uudet termit kasvattivat eri termikaavojen mukaisten termien lukumäärä, mutta suhteelliset osuudet pysyivät lähes samoina. En ole luetellut harvinaisia termikaavoja, koska tutkimukseni tarkoitus on tarkastella eri menetelmillä poimittujen termien ja termiehdokkaiden tyypillisiä ominaisuuksia, ei kaikkia mahdollisia vivahteita.

TAULUKKO 6. Suomalaisten asiantuntijoiden poimimien termien yleisimmät termikaavat

Termikaava	1. kierros f	%	Valmis f	termilista %	Esimerkki
N <sub>c</sub>	42	35,6	59	31,2	laivaliikenne
Abb-N, Abb N	22	18,7	24	12,7	VTS operaattori
N	21	17,8	35	18,5	luotsi
N <sub>(c)</sub> N <sub>(c)</sub>	13	11,0	23	12,2	laivaliikenteen ohjaus
A N <sub>(c)</sub>	9	7,6	16	8,5	elektroninen tietojenvaihto
Abb	6	5,1	10	5,3	VTS
N <sub>(c)</sub> N C N <sub>c</sub>	3	2,5	4	2,1	meriliikenteen ohjaus- ja informaatiopalvelu
en (X) <sup>a</sup>	2	1,7	12	6,3	Geographical Information System
Yhteensä	118	100	183	96,8	

Huom. Verbien partisippiimuodot substantiivien määritteinä on luokiteltu adjektiiveiksi.

<sup>a</sup> X = substantiivilauseke

On hämmästyttävää, että verbitermit puuttuvat lähes kokonaan suomalaisten asiantuntijoiden valitsemien termien listasta, sillä esimerkiksi heti yksimielisesti valituista 118 termistä noin kolmannes eli 41 termiä nimeää toimintakäsitettä (*laivaliikenteen ohjaus, navigointi*). Tästä huolimatta lopullisessa termilistassa on vain yksi verbitermi, *ohjailla*. Termiä arvioimiseen ei termikaavoista kuitenkaan ole apua, sillä termikaavoissa ei juuri ole eroja yksimielisesti poimittujen ja neuvottelun jälkeen valmiiseen termilistaan hyväksytyjen termien välillä.

Samaa käsitettä nimeävistä termivarianteista on suomalaisten asiantuntijoiden termilistassa runsaasti esimerkkejä. Suomalaisten asiantuntijoiden valitsemissa termeissä variaatiota esiintyy varsinkin sanaliitoissa, jotka nimeävät tai kuvaavat erikoisalan (uusia) käsitteitä. Tähän ryhmään kuuluville termeille tyypillistä on vakiintumattomuus, joka ilmenee runsaana morfologisena vaihteluna. Asiantuntijoiden termilistassa on 27 sanaliittotermiä, jos *VTMIS*- tai *VTS*-lyhenteen kanssa ilman yhdysviivaa muodostetut termit jätetään huomiotta. Sanaliittotermeistä noin puolet esiintyy termilistassa ainakin kahdessa erilaisessa muodossa, mutta saman käsitteen nimityksenä. Esimerkiksi sanaliittotermi *meriliikenteen ohjausjärjestelmä* on asiantuntijoiden termilistassa myös yhdyssanana *meriliikenteenohjausjärjestelmä*. Myös sanaliittotermi *tietojen vaihto* on yhdyssanana termeissä *elektroninen tietojenvaihto*, *horisontaalinen tietojenvaihto* ja *vertikaalinen tietojenvaihto*. Nämä termit saattavat olla esimerkki Salmen (2003: 12) mainitsemasta erosta termin ja yleiskielen ilmaisun välillä. Salmen mukaan sanaliiton sanat muodostavat termin yhteen kirjoitettuna, mutta erikseen kirjoitettuna ne ovat yleiskieltä (ks. jakso 5.4).

Suomalaisten asiantuntijoiden termilistassa on viisi esimerkkiä syntaktisesta variaatiosta. Asiantuntijoiden termilistassa ovat termit *integroitu tutkaverkko*, *kokonaiskuljetusketju*, *tutka-asemien verkko*, *ympäristönsuojeluviranomainen* ja *sisääntuloväylä*. Samoja käsitteitä nimetään tekstissä myös näiden termien syntaktisilla varianteilla *integroitu verkko*, *kuljetusketju*, *tutkaverkko*, *ympäristöviranomainen* ja *sisääntulo*. Nämä termivariantit voisivat olla esimerkkejä myös ellipsisistä. Olen kuitenkin luokitellut ne esimerkeiksi syntaktisesta variaatiosta, koska variantit ovat termin muunnoksia, joissa



termistä on jätetty jokin osa pois (ks. jakso 5.4). Mielenkiintoista on, että poisjäävä osa voi olla termin alussa, keskellä tai lopussa.

Semanttisessa variaatiossa jokin termin osa on korvattu synonyymilla. Suomalaisten asiantuntijoiden termilistassa on esimerkkejä myös tästä variaatiolajista. Esimerkiksi termeissä *meriliikenteen ohjaus- ja informaatiopalvelu* ja *laivaliikenteen ohjaus- ja informaatiopalvelu* sanaliiton ensimmäinen osa on korvattu tekstin kontekstissa synonyymisella sanalla. Mielenkiintoisia ovat termit *yhteistoimintakomitea*, *yhteistyökomitea* ja *yhteistyötoimikunta*, jotka ilmeisesti nimeävät samaa käsitettä. Kahdessa ensimmäisessä termivariantissa alkuosat *yhteistoiminta* ja *yhteistyö* ovat synonyymisia. Kahdessa jälkimmäisessä taas loppuosat *komitea* ja *toimikunta* ovat synonyymisia.

Asiantuntijatkin ovat joskus erimielisiä siitä, mitkä määritteet kuuluvat termiin. Määritteiden pois jättäminen tuottaa elliptisiä termejä. Ellipsi esiintyy silloin, kun käsitteeseen viitataan, mutta sitä ei ole eksplisiittisesti tekstissä, vaan tekstissä on esimerkiksi yläkäsitteen nimitys. Ilmiö on yleinen erikoisalateksteissä kielestä riippumatta (ks. esim. Bowker & Pearson 2002: 169). Suomalaisten asiantuntijoiden termilistasta poimittuja esimerkkejä elliptisistä termeistä ovat termit *tunnistaminen*, *sijainti*, *turvallisuus*, *informaatiopalvelu*, *ohjausjärjestelmä* ja *päällikkö*, joista kaikista termilistassa on myös termin pitempi muoto: *laivojen tunnistaminen*, *laivan sijainti*, *liikenneturvallisuus* ja *meriliikenteen turvallisuus*, *liikenteen informaatiopalvelu* ja *aluksen päällikkö*. Elliptiselle termille *ohjausjärjestelmä* termilistassa on neljä pitempää varianttia: *liikenteen ohjausjärjestelmä*, *liikenteenohjausjärjestelmä*, *meriliikenteen ohjausjärjestelmä* ja *meriliikenteenohjausjärjestelmä*. Viimeksi mainittu termivariantti on todiste suomen kielen agglutinoivasta luonteesta, onhan siinä puristettu yhteen sanaan termi, jonka vastine esimerkiksi venäjän kielessä on neljän sanan mittainen (*система управления движением судов*).

Elliptisissä ilmauksissa on usein mukana konjunktio *ja*. Terminpöimintakokeen suomenkielisessä tekstissä on esimerkiksi seuraavat *ja*-konjunktin sisältävät ilmaukset, joissa on vähintään kaksi termiä ja osa termeistä on elliptisiä: *laivojen ilmaiseminen*, *tunnistaminen ja jäljittäminen* (*laivojen ilmaiseminen*, *laivojen tunnistaminen*, *laivojen jäljittäminen*), ja *etsintä-*, *pelastus-* ja *ympäristönsuojeluviranomaiset* (*etsintäviranomaiset*, *pelastusviranomaiset*, *ympäristönsuojeluviranomaiset*).

Suomalaisten asiantuntijoiden termilistassa on huomiota herättävän runsaasti englanninkielisiä termejä ja lyhenteitä. Koska moni niistä nimeää samoja käsitteitä kuin suomenkieliset termit, käsittelen niitä tässä yhteydessä. Listassa on 11 kokonaan englanninkielistä monisanaista termiä ja 10 englannin sanoista muodostettua lyhennettä. Asiantuntijat valitsivat englanninkieliset termit listaan, vaikka niistä kahdeksalle listassa on myös suomenkielinen termivastine, joka nimeää samaa käsitettä kuin englanninkielinen termi. Vastaavasti listassa on suomenkielinen termivastine kuudelle asiantuntijoiden valitsemalle lyhenteelle. Esimerkiksi termille *Regional Traffic Information Service* ja lyhenteelle *RTIS* on vastine *alueellinen liikenteen informaatiopalvelu*. Kiintoisan termiryppään muodostavat lyhennetermit *VTS* ja *VTMIS* sekä niiden pitkät muodot *Vessel Traffic Service* sekä *Vessel Traffic Management and Information Service* ja *Vessel Traffic Management and Information System*. Termille *Vessel Traffic Service* on listassa vastineet *laivaliikenteen ohjaus*, *liikenteen ohjaus*, *liikenteenohjausjärjestelmä*, *meriliikenteen ohjausjärjestelmä*, *meriliikenteenohjausjärjestelmä* ja *ohjausjärjestelmä*. Termille *Vessel Traffic Management and Information System*, joka on Google-hakukoneen antaman tuloksen perusteella termin tavallisempi muoto, on listassa vastineet *laivaliikenteen ohjaus- ja informaatiopalvelu*, *laivaliikenteen ohjaus- ja informaatiojärjestelmä* sekä *meriliikenteen ohjaus- ja informaatiopalvelu*. Vastineiden runsaus johtuu käsitteen suomenkielisen nimityksen vakiintumattomuudesta.

Suurin frekvenssi tekstissä on lyhenteillä *VTS* ja *VTMIS*, mutta termeiksi on valittu myös esimerkiksi lyhenteet *IALA*, *IMCO* ja *IMO*, jotka ovat tietyistä merenkulun järjestöistä käytettyjä lyhenteitä. Valinta on yllättävä, sillä organisaatiot ovat yksilökäsitteitä. Niiden nimitykset eivät siis täytä termin määritelmän ehtoa, jonka mukaan termi nimeää erikoisalan yleiskäsitettä. Lisäksi listassa on 24 sellaista termiä, joissa englanninkielinen lyhenne on termin osa. Englanninkielisen lyhenteen ja suomenkielisen sanan yhdistelmätermien määrää kasvattaa lyhenne *VTS*, joka on termin osana 16 eri termissä, kuten esimerkiksi termeissä *VTS palvelu* ja *VTS-verkko*. Kaikki termilistan lyhenteet, joko itsenäisinä termeinä tai termin osina, on muodostettu englanninkielisistä sanoista. Niistä on tullut suomenkielisiä termejä joko sellaisenaan, jolloin niihin liitetään tekstissä suomen sijapäätteet, esimerkiksi *VTS:ssä*, tai termin osina, jolloin niihin liitetään suomenkielinen sana joko yhdysmerkillä tai ilman, esimerkiksi *EEZ/LRT-järjestelmä* ja *off-shore VTS toiminta*. Englanninkieliset lyhenteet, niiden englanninkieliset avatut muodot ja suomenkieliset vastineet kulkevat merenkulun teksteissä termeinä rinta rinnan ainakin tämän terminpoimintakokeen perusteella.

Terminpoimintakokeeni suomenkielisessä aineistossa variaatio on siis yleistä. Suomalaisten asiantuntijoiden kokoamassa 189 termin listassa on yli 20 variaatioesimerkkiä, jotka edustavat kuutta erilaista variaatiolajia. Joskus varianttien erot ovat varsin pieniä, kuten varianteilla *VTS operaattori* ja *VTS-operaattori*. Ilmeisesti asiantuntijat ovat halunneet antaa kirjoittajalle anteeksi pienet poikkeamat suomen kielen vakiintuneesta kirjoitustavasta. Varianttien valinta termilistaan osoittaa, että asiantuntijat poimivat termejä ennemmin käsitelähtöisesti kuin termilähtöisesti. Koeasetelmassa ei kuitenkaan ole mahdollista päätellä, ovatko asiantuntijat ja koehenkilöt tunnistaneet kaikki termivariantit samojen käsitteiden nimityksiksi. Sisältääkö valittu suomenkielinen teksti poikkeuksellisen paljon termivariaatiota vai onko termivariaatio todella näin yleistä?

### 6.3.2 Venäläisten asiantuntijoiden valitsemien termien ominaisuuksia

Venäläisten asiantuntijoiden valitsemien termien määrä on huomattavasti pienempi kuin suomalaisten asiantuntijoiden valitsemien. Syynä saattaa olla yhtä valintakierrosta lyhyempi valintaprosessi, tekstin sisältämien termien pienempi määrä tai yksilölliset erot. Valintaprosessin ja tuloksen eroista huolimatta tarkastelen venäläisten asiantuntijoiden valitsemista termeistä samoja piirteitä kuin suomalaisten asiantuntijoiden valitsemista. Selostan ensin termien pituutta sanoina. Sitten tarkastelen termien frekvenssiä ja sanaluokkia termikaavoina ilmaistuna. Jakson lopussa tarkastelen lyhyesti venäläisten asiantuntijoiden valitsemien termien vierasperäisyyttä ja variaatiota.

Venäläisten asiantuntijoiden heti ensimmäisellä kierroksella yksimielisesti valitsemien termiehdokkaiden pituus on keskimäärin 1,65 sanaa. Joukossa on kaksi kolmen sanan ja yksi neljän sanan mittainen termi. Yksi- ja kaksisanaisten termien osuudet ovat täsmälleen samansuuruiset eli vähän alle 46 prosenttia. Sanaksi on laskettu merkkijono, jonka kummallakin puolella on välilyönti. Lyhennetermit on laskettu yhdeksi sanaksi, ja prepositiot on otettu mukaan sanalukuun. Kaikissa tutkimuksissa ei prepositioita ole laskettu erillisiksi sanoiksi, koska niillä ei yksinään ole omaa merkitystä (Daille ym. 1996: 208). Käyttämälläni laskentatavalla pisimpien venäläisten termien pituus kasvaa yhdellä tai kahdella sanalla verrattuna laskentatapaan, jossa prepositioita ei lasketa sanoiksi.

Venäläisten asiantuntijoiden lopullisessa termilistassa keskimääräinen terminpituus on 2,36 sanaa. Kolmisanaisten ja vähintään nelisanaisten termien osuus onkin noussut nelinkertaiseksi eli 21,5 ja 15,0 prosenttiin. Neuvottelua tarvittiin siis erityisesti monisanaisten termiehdokkaiden termiästä päättämiseen.

TAULUKKO 7. Venäläisten asiantuntijoiden poimimien termien pituus sanoina

Sanaluku	1. valintakierros (A=1,65)		Termilista (A=2,36)	
	f	%	f	%
1	17	45,9	28	26,1
2	17	45,9	40	37,4
3	2	5,5	23	21,5
≥ 4	1	2,7	16	15,0
Yhteensä	37	100	107	100

Suomalaisten asiantuntijoiden tavoin venäläistenkin asiantuntijoiden näkemykset määriteosien kuulumisesta termiin vaihtelivat. Yhden venäläisen asiantuntijan valitseman termiehdokkaan alussa saattoi olla määriteosa, jota ei toisten valitsemisissa termiehdokkaissa ollut (*государственный контроль над морским транспортом, локальная ЕСКУС, основной порт*). Termilistaan jäivät lopulta lyhyemmät muodot *контроль над морским транспортом, ЕСКУС ja порт*.

Myös termiehdokkaan lopussa saattoi olla ylimääräistä kielellistä ainesta. Venäläisissä termiehdokkaissa määriteosa on usein genetiivimuotoinen substantiivi, joka on kielen syntaktisesta rakenteesta johtuen pääsanana jäljessä. Asiantuntijat eivät aina olleet yksimielisiä genetiivimääritteiden kuulumisesta termiin (*вектор скорости участвующих судов, внутренние территориальные воды залива*). Määritteet jätettiin lopulta pois termistä.

Venäläisten asiantuntijoiden heti yksimielisesti valitsemien termien frekvenssi on keskimäärin 5,40 mutta valmiin termilistan termien frekvenssi on vain 3,95. Osittain ero selittyy muutaman yleisesti tekstissä esiintyvän termin frekvenssillä. Yksimielisesti ensimmäisellä kierroksella valittujen termien joukossa on sellaisia lyhenteitä ja kaikilla merenkulun aloilla esiintyvien käsitteiden nimityksiä, joilla on tekstissä suuri frekvenssi. Esimerkiksi *АИС* (AIS), *ССС* (alusilmoitusjärjestelmä), *СУДС* (VTS), *порт* (sattama), *судно* (alus), *судоходство* (alusliikenne) esiintyvät tekstissä vähintään 15 kertaa. Toisaalta termit *ЕСКУС* (VTMIS), *район А1* (alue A1) ja *бассейн* (vesistöalue), jotka esiintyvät lähtötekstissä 49, 18 ja 16 kertaa, ovat tulleet mukaan vasta lopulliseen listaan. Korkeintaan kerran tekstissä esiintyvien termien osuus on suuri jo ensimmäisellä kierroksella valituista termeistä, mutta lopullisen termilistan termeistä niiden osuus on jo yli puolet, 52,3 prosenttia. Neuvottelun jälkeen termilistaan hyväksyttiin erityisesti kerran tekstissä esiintyviä termiehdokkaita. Samoin kuin suomalaisten asiantuntijoiden terminpoiminnassa venäläisten asiantuntijoiden terminpoiminnassa esiintymien määrä tekstissä selittää sitä, millä perusteella jotkin termit valitaan heti yksimielisesti ja jotkin vasta neuvottelun jälkeen.

TAULUKKO 8. Venäläisten asiantuntijoiden poimimien termien frekvenssi

Frekvenssi	1. valintakierros (A=5,40)		Termilista (A=3,95)	
	f	%	f	%
≤ 1	15	40,6	56	52,3
2	4	10,8	14	13,1
3–4	7	18,9	16	15,0
≥ 5	11	29,7	21	19,6
Yhteensä	37	100	107	100

Suomalaisten asiantuntijoiden tavoin venäläiset asiantuntijat olivat hyvin yksimielisiä termien termikaavoista. Kaikki venäläisestä tekstistä poimitut ja yhteiseen termilistaan valitut termit ovat substantiivilausekkeita. Tosin yksi asiantuntija valitsi termiehdokkaiksi myös verbejä, adjektiiveja ja vakiintuneita fraaseja, mutta toiset valintaan osallistuneet eivät vakuuttuneet niiden termiydestä.

Kaikki ensimmäisellä kierroksella valitut termit sopivat kuuteen termikaavaan. Liki kolmannes eli 29,7 prosenttia venäläisten asiantuntijoiden heti yksimielisesti valitsemista termeistä on adjektiivin ja substantiivin muodostamia sanaliittoja eli noudattavat termikaavaa A N, esimerkiksi *подходный путь* (tuloväylä). Lähes yhtä paljon eli 27 prosenttia on yksisanaista substantiiveja N<sup>1</sup>, esimerkiksi *судоходство*. Seuraavaksi suurin ryhmä ovat lyhennetermit Abb, esimerkiksi *СУДС*, joita on listassa 18,9 prosenttia. Muut termit ovat kahden, kolmen tai neljän substantiivin sanaliittoja tai adjektiivin ja substantiivin sanaliittoja, joissa adjektiiveja tai substantiiveja on enemmän kuin yksi, esimerkiksi *безопасность мореплавания* (merenkulun turvallisuus), *зона разделения движения* (liikenteenjakovyöhyke), *система управления движением судов* (alusliikennepalvelu), *спутниковая система связи* (satelliittiviestintäjärjestelmä). Näissä termeissä on usein vähintään termin osana sana *море* (meri), *порт* (satama), *судно* (alus) tai näistä muodostettu adjektiivi.

Suomalaisten asiantuntijoiden tavoin venäläisetkin asiantuntijat olivat valinneet yksilökäsitteiden nimiä termiehdokkaiksi. Organisaatioiden ja asiakirjojen nimet (*Минтранс РФ*, *Парижский меморандум*) sekä niiden nimistä muodostetut lyhenteet (*ИМО*, *МКУБ*) ja muut yksilökäsitteiden nimet (*ГМССБ*, *Мировой океан*) jakoivat kuitenkin venäläisten asiantuntijoiden mielipiteet. Toisen kierroksen jälkeen ne karsuivatkin termilistasta. Lyhenteitä jäi venäläisten asiantuntijoidenkin termilistaan, mutta ne ovat kolmea englanninkielistä lyhennettä lukuun ottamatta pitkien (vähintään kolmisanaisten) termien osista muodostettuja.

Aiemman tutkimuksen mukaan venäläisten termien yleisimmät termikaavat ovat partisiipin ja substantiivin yhdistelmä, kahden substantiivin ja niitä määrittävän adjektiivin yhdistelmä (A (N N)) sekä adjektiivin, substantiivin ja niitä määrittävän adverbin yhdistelmä (Ad (A N)) (Citkina 1988: 103). Tutkimukseni tulos on samansuuntainen kuin Citkinan tulos, mutta terminpoimintakokeessani yleisin venäläinen termikaava on adjektiivin ja yhden substantiivin yhdistelmä A N, sillä olen luokitellut verbin partisiippimuodot adjektiiveiksi. Toiseksi eniten on yhden substantiivin muodostamia yksisanaista termejä.

Lopulliseen listaan valitut venäläiset termit eivät keskity tiettyihin termikaavoihin niin selvästi kuin suomalaiset, sillä termikaavoja on listassa kaikkiaan 25 ja kuuteen tavallisimpaan kaavaan sopii 72,9 prosenttia termeistä. Erilaisten termikaavojen määrä on huomattavan suuri, koska uudet termit ovat monisanaista. Ensimmäisellä kierroksella

<sup>1</sup> Venäjänkielisten termien termikaavoissa ei yhdyssubstantiiveja ole eroteltu muista substantiiveista.

valittujen termien listassa ei ole yhtään termiä, jonka osana olisi konjunktio tai prepositio. Lopullisessa termilistassa sen sijaan on 18 termiä, jossa on vähintään yksi konjunktio, prepositio tai symboli (*единая система контроля и управления судоходством, система по надзору за судоходством, морской район А1*). Koska uudet termit kasvattivat erilaisten termikaavojen lukumäärää, yleisimpien termikaavojen mukaisten termien suhteelliset osuudet laskivat muutaman prosentin.

TAULUKKO 9. Venäläisten asiantuntijoiden poimimien termien yleisimmät termikaavat

Termikaava	1. kierros		Valmis termilista		Esimerkki
	f	%	f	%	
A N	11	29,8	22	20,6	подходный путь
N	10	27,0	16	15,0	судоходство
N N	6	16,2	13	12,1	безопасность мореплавания
Abb	7	18,9	12	11,2	СУДС
A A N	0	0,0	7	6,5	морская спутниковая связь
N N N	1	2,7	5	4,7	зона разделения движения
A <sub>x</sub> N <sub>x</sub> tai N <sub>x</sub>	2	5,4	8	7,5	система управления движением судов
Yhteensä	37	100	83	77,6	

Huom. A<sub>x</sub>N<sub>x</sub> = vähintään kolmisanainen termi, jossa on n-kappaletta adjektiiveja ja substantiiveja ja jossa n = 1–4, N<sub>x</sub> = n-kappaletta substantiiveja, jossa n = ≥ 4

Suomalaiset asiantuntijat valitsivat termilistaansa yhden verbitermin, mutta venäläiset eivät ainoatakaan. Venäläisten termien termiydestä voidaan termikaavojen perusteella sanoa, että jos termiehdokas noudattaa kaavaa A<sub>x</sub> N<sub>x</sub> tai N<sub>x</sub>, jossa adjektiivien ja substantiivien määrä voi vaihdella yhdestä neljään, termiehdokas on hyvin todennäköisesti termi.

Suomenkielisen koetekstin tavoin venäjänkielinen koeteksti sisältää termivariaatiota ja elliptisiä ilmauksia, jotka aiheuttivat eroja termiehdokaslistoihin. Esimerkiksi termiehdokkaat *контроль за судами* ja *контроль над судами* eivät lopulta päätyneet asiantuntijoiden termilistaan, jonne valittiin virallisemmän kuuloiset *контроль за движением судов* ja *контроль над судоходством*. Termin kaltainen kielellinen yksikkö *подход к порту* jäi pois termilistasta, mutta samaa käsitettä nimeävä *подходный путь* valittiin termiksi yksimielisesti.

Venäläisten termien listassa ei ole samanlaista ortografista variaatiota kuin suomalaisten termien listassa, sillä termeillä ei ole variantteja, joissa terminosat olisi kirjoitettu yhteen tai erikseen, yhdysmerkillä tai sitä ilman. Vaihtelua ei ole myöskään isojen tai pienten kirjainten välillä. Sen sijaan syntaktisesta variaatiosta on venäläistenkin termien listassa esimerkkejä. Termille *бассейновая информационная система* on termilistassa variantti *бассейновая система*, jossa sana *информационная* puuttuu, ja termille *контроль движения судов* variantti *контроль за движением судов*, jossa substantiivin *движение* sijamuoto on vaihtunut. Venäläisessä termilistassa on myös esimerkki morfosyntaktisesta variaatiosta, jossa yhden terminosan sanaluokka vaihtuu. Termille *обеспечение безопасного плавания* on variantti *обеспечение безопасности мореплавания*. Koska varianteissa yksi terminosa kuuluu eri sanaluokkaan, on varianteilla myös erilainen termikaava, N A N tai N N N. Semanttisesta variaatiosta on termilistassa yksi



esimerkki. Termeissä *контроль над морским транспортом* ja *контроль над судоходством* on yksi terminosa korvattu synonyymisella ilmauksella, sillä tekstin kontekstissa sanaliitto *морской транспорт* (merikuljetus) ja substantiivi *судоходство* (alusliikenne) ovat synonyymisia. Termilistassa on kolmaskin termi, joka nimeää samaa käsitettä, sillä termissä *надзор за судоходством* ensimmäinen osa *надзор* (valvonta) on synonyyminen termin *контроль над судоходством* ensimmäisen osan *контроль* (valvonta) kanssa. Suomalaisia käsitteitä *alusliikennepalvelu* ja *alusliikenteen ohjaus- ja informaatiopalvelut* vastaavien venäläisten käsitteiden nimityksissä ei ole variaatiota, koska nimitykset *система управления движением судов* ja *единая система контроля и управления судоходством* esiintyvät tekstissä avattuna vain kerran ja muulloin kirjoittajat käyttävät lyhenteitä *СУДС* ja *ЕСКУС*.

Terminpoimintakokeen venäläisessä koetekstissä ja termilistassa on esimerkkejä elliptisistä ilmauksista, joissa adjektiivimäärite *морской* (meri-) on jätetty pois termistä. Tekstissä saatetaan puhua esimerkiksi käsitteistä *морской бассейн*, *морской район А1* tai *морской флот*. Niihin viitataan kuitenkin usein elliptisellä ilmauksella *бассейн*, *район А1* ja *флот*. Asiantuntijat ovatkin valinneet termilistaan sekä pitkät että elliptiset termimuodot.

Venäläisten asiantuntijoiden kokoamassa termilistassa ei ole yhtään englanninkielistä avattua termiä. Venäläisten ja suomalaisten asiantuntijoiden välinen ero englanninkielisten termien valinnassa selittyy osittain tekstien eroilla, sillä venäjänkielisessä artikkelissa *Eskos* on vain yksi englanninkielisen termin avattu muoto: *Vessel Traffic Management and Information System*, jota asiantuntijat eivät siis valinneet. Sen sijaan he valitsivat kaksi kertaa tekstissä esiintyvän englanninkielisen lyhennetermin *VTMIS* ja sen venäjänkielisen vastineen *единая система контроля и управления судоходством*. Myös muut tekstin englanninkieliset lyhennetermit, *DGPS* ja *GPS*, on valittu termeiksi, vaikka niillä kummallakin on lähtötekstissä vain yksi esiintymä. Englanninkielisten lyhenteiden esiintymien vähäisyys ei kuitenkaan tarkoita, että venäjänkielisessä tekstissä ei esiintyisi lyhenteitä, sillä tekstissä on 29 venäjänkielistä lyhennettä, ja niillä yhteensä 210 esiintymää. Lyhenteistä yhdeksän on asiantuntijoiden valinnan mukaan termejä. Lisäksi asiantuntijat valitsivat kolme termiä, joissa lyhenne on termin osana. Venäläisen lyhennetermin ja termin pitkän muodon välisestä variaatiosta on monta esimerkkiä, sillä kaikki venäläiset lyhennetermit *АИС*, *ЕСКУС*, *МАП*, *СКП*, *СКЦ*, *СПЦ*, *ССС*, *СУДС* ja *ЦМСС* ovat listassa myös avattuina venäjänkielisinä termeinä (*автоматическая идентификационная система*, *единая система контроля и управления судоходством*, *морская администрация порта*, *служба капитана порта*, *спасательно-координационный центр*, *спасательно-координационный подцентр*, *система судовых сообщений*, *система управления движением судов* ja *центр морской спутниковой связи*). Venäläisessäkin termilistassa on siis termivariaatiota, mutta ei niin paljon kuin suomalaisessa termilistassa.

#### 6.4 Suomalaisen opiskelijaryhmien terminpoiminta

Tutkimukseeni kuuluvassa terminpoimintakokeessa yhtenä tavoitteena oli selvittää, miten hyvin koulutustalustaan erilaisten koehenkilöryhmien terminpoiminta vastaa asiantuntijoiden tekemää terminpoimintaa. Tavoitteena oli myös vastata kysymykseen, onko koehenkilöryhmien terminpoiminnan tuloksessa eroja, jotka selittyvät toisaalta koehenkilöiden hyvällä tai huonolla erikoisan tuntemuksella tai toisaalta hyvällä tai huonolla terminologian teorian tuntemuksella. Tähän kysymykseen etsin vastausta vertaamalla koehenkilöryhmien terminvalintaa asiantuntijoiden valintaan ryhmittäin.



Koehenkilöryhmien ja asiantuntijoiden terminvalinnan vertailua varten muodostin kummankin ryhmän termiehdokaslistoista yhteislistan, jossa ovat vähintään neljän samaan ryhmään kuuluvan koehenkilön valitsevat termiehdokkaat. Kääntämisen opiskelijoiden termiehdokaslistoissa on 238 ja merenkulun opiskelijoiden listoissa 191 tämän kriteerin täyttävää termiehdokasta. Koska yli yksi kolmasosa ryhmän koehenkilöistä on pitänyt näitä termeinä, niitä voidaan pitää jo erittäin hyvinä termiehdokkaina. Lisäksi vähintään neljän koehenkilön valitsemien termiehdokkaiden lukumäärä on lähellä ryhmän keskiarvoa, joka on 224 kääntämisen opiskelijoiden ryhmässä ja 180 merenkulun opiskelijoiden ryhmässä.

Vertailua varten kokosin ryhmittäin suppeammat listat, joissa ovat kunkin opiskelijaryhmän ja asiantuntijoiden yksimielisesti valitsevat termit, kunkin opiskelijaryhmän listasta puuttuvat mutta asiantuntijoiden valitsevat termit ja kunkin opiskelijaryhmän valitsevat mutta asiantuntijoiden hylkäämät termiehdokkaat. Opiskelijoiden listasta puuttuvista termeistä käytän tässä vertailussa nimitystä *piilotermit* ja opiskelijoiden valitsemista ylimääräisistä termiehdokkaista nimitystä *valetermi*.

Seuraavissa jaksoissa selostan vertailujen tuloksia. Ensin tarkastelen termejä, jotka ovat samoja sekä vertailussa olevassa termiehdokaslistassa että asiantuntijalistassa. Toiseksi tarkastelen vertailtavan koehenkilöryhmän termiehdokaslistan piilotermejä. Kolmanneksi tarkastelen termiehdokaslistan valetermejä. Kaikissa näissä vaiheissa tarkastelen termien tai termiehdokkaiden pituutta sanaluvulla mitattuna, frekvenssiä lähtötekstissä ja termin tai termiehdokkaan osien sanaluokkia eli termikaavoja. Poiminnan onnistumisen mittareina käytän saantia ja tarkkuutta. Suomalaisten kääntämisen opiskelijoiden ryhmästä käytän vertailuissa lyhennettä *kofi* ja merenkulun opiskelijoiden ryhmästä lyhennettä *mofi* (ks. jakso 6.1).

#### **6.4.1 Suomalaisten opiskelijaryhmien poimimien termien ominaisuuksia**

Seuraavaksi tarkastelen niitä kustakin opiskelijaryhmästä vähintään neljän koehenkilön valitsemia termiehdokkaita, jotka myös asiantuntijat olivat valinneet ja joita siis voi nimittää termeiksi. Tarkasteltavia piirteitä ovat termien pituus sanaluvulla mitattuna, frekvenssi ja termikaavat. Tavoitteena on vastata kysymyksiin, millaisia ovat termit, jotka kääntämisen opiskelijat valitsivat, millaisia ovat termit, jotka merenkulun opiskelijat valitsivat, ja eroavatko koehenkilöryhmien valitsevat termit sanaluvultaan, frekvenssiltään tai termikaavoiltaan.

Suomalaisten koehenkilöiden valitsemien termien pituus sanaluvulla mitattuna ei eroa merkittävästi asiantuntijoiden valitsemien termien pituudesta. Vähintään neljän suomalaisen kääntämisen opiskelijan valitsemien termien keskimääräinen sanaluku on 1,62 eli hieman korkeampi kuin suomalaisten asiantuntijoiden valitsemien termien keskimääräinen sanaluku, joka on 1,54. Merenkulun opiskelijoiden valitsemien termien keskimääräinen sanaluku sen sijaan on aavistuksen alhaisempi, 1,51. Kääntämisen opiskelijoiden terminvalinta onnistui sitä paremmin mitä pitempiä poimittavat termit ovat. Asiantuntijoiden valitsemista yksisanaisista termeistä kääntämisen opiskelijat valitsivat 69 prosenttia ja kaksisanaisista 74,1 prosenttia. Kolmisananaisista termeistä saanti on jo erittäin hyvä, 83,3 prosenttia, ja vähintään nelisanaisista termeistä 100 prosenttia. Merenkulun opiskelijat sen sijaan valitsivat keskimäärin enemmän yksisanaisia ja vähemmän kaksi- tai kolmisananaisia termejä kuin asiantuntijat. Merenkulun opiskelijoiden ryhmässä yksisanaisien termien saanti on 72,4 prosenttia, mutta laskee 57,4 prosenttiin kaksisanaisien ja 58,3 prosenttiin kolmisananaisissa termeissä. Vähintään nelisanaisista termeistä puuttuu yksi, jolloin saanti on 85,7 prosenttia.

TAULUKKO 10. Suomalaisten asiantuntijoiden ja vähintään neljän koehenkilön poimimien termien pituus

Sanaluku	atfi (A=1,54)		kofi (A=1,62)		mofi (A=1,51)	
	f	%	f	%	f	%
1	116	61,4	80	58,4	84	65,6
2	54	28,6	40	29,2	31	24,2
3	12	6,3	10	7,3	7	5,5
≥ 4	7	3,7	7	5,1	6	4,7
Yhteensä	189	100	137	100	128	100

Suurin osa eli 58,4 prosenttia kääntämisen opiskelijoiden ja 65,6 prosenttia merenkulun opiskelijoiden termeistä on yhden sanan mittaisia ja korkeintaan kahden sanan termejä on kummassakin ryhmässä lähes 90 prosenttia. Nämäkin osuudet ovat hyvin lähellä samat kuin asiantuntijoiden termilistassa. Silmiinpistävää on, että vähintään kolmen sanan mittaisista termeistä noin puolet on englanninkielisiä sekä asiantuntijoiden että opiskelijoiden listoissa. Todellisuudessa vähintään kolmisanaisten suomalaisten termien osuus on siis vain noin viisi prosenttia.

Koehenkilöiden valitsemien termien frekvenssi ei sekään eroa oleellisesti suomalaisten asiantuntijoiden valinnasta. Kääntämisen opiskelijoiden ryhmässä keskimääräinen frekvenssi on 2,58 ja merenkulun opiskelijoiden ryhmässä 2,94, kun se asiantuntijoiden valinnassa on 2,34. Yli puolet eli 62,1 prosenttia kääntämisen opiskelijoiden ja 59,4 prosenttia merenkulun opiskelijoiden valitsemista termeistä esiintyy lähtötekstissä korkeintaan kerran. Asiantuntijoiden valinnassa pienifrekvenssisten ( $f \leq 1$ ) termien osuus on vieläkin suurempi, 65,6 prosenttia, minkä vuoksi koehenkilöiden valitsemien termien keskimääräinen frekvenssi on hieman korkeampi kuin asiantuntijoiden valitsemien.

TAULUKKO 11. Suomalaisten asiantuntijoiden ja vähintään neljän koehenkilön poimimien termien frekvenssi

Frekvenssi	atfi (A=2,34)		kofi (A=2,58)		mofi (A=2,94)	
	f	%	f	%	f	%
≤ 1	124	65,6	85	62,1	76	59,4
2	35	18,5	28	20,4	24	18,7
3	6	3,2	8	5,8	6	4,7
≥ 4	24	12,7	16	11,7	22	17,2
Yhteensä	189	100	137	100	128	100

Noin viidennes kummankin ryhmän valitsemista termeistä esiintyy lähtötekstissä kaksi kertaa. Ero koehenkilöryhmien välillä näkyy myös suurifrekvenssisten termien valinnassa, sillä kääntämisen opiskelijoiden valitsemista termeistä vain 11,7 prosenttia on suurifrekvenssisia termejä eli termejä jotka esiintyvät tekstissä vähintään neljä kertaa. Merenkulun opiskelijoiden ryhmässä näiden termien osuus on 17,2 prosenttia. Asiantuntijoiden valitsemista termeistä 12,7 prosenttia esiintyy lähtötekstissä vähintään neljä kertaa.

Termikaavaltaan vähintään neljän suomalaisen koehenkilön valitsemat termit eivät eroa oleellisesti suomalaisten asiantuntijoiden valinnasta. Yleisin termikaava kummasakin ryhmässä on yhdyssubstantiivi  $N_c$ , joita termeistä on kääntämisen opiskelijoiden

ryhmässä 31,4 prosenttia ja merenkulun opiskelijoiden ryhmässä 33,6 prosenttia. Adjektiivin ja substantiivin tai kahden substantiivin sanaliittotermien osuus jää kääntämisen opiskelijoiden ryhmässä 19:ään ja merenkulun opiskelijoiden ryhmässä 14 prosenttiin. Yllättävää on, että suomenkielisestä lähtötekstistä valituista termeistä noin 30 prosenttia kummassakin ryhmässä on englanninkielisiä sanaliittotermiä, englanninkielisiä lyhenteitä tai englanninkielisen lyhenteen ja suomenkielisen sanan yhdistelmiä, kuten termit *Geographical Information System, VTS* ja *VTMIS-järjestelmä*.

TAULUKKO 12. Suomalaisten asiantuntijoiden ja vähintään neljän koehenkilön poimimien termien tavallisimmat termikaavat

Termikaava	atfi		kofi		mofi	
	f	%	f	%	f	%
N <sub>c</sub>	59	31,2	43	31,4	43	33,6
Abb N	23	12,2	23	16,8	19	14,8
N	35	18,5	17	12,4	23	18,0
N <sub>(c)</sub> N <sub>(c)</sub>	23	12,2	13	9,5	13	10,2
A N <sub>(c)</sub>	16	8,5	13	9,5	5	3,9
Abb	11	5,8	9	6,5	10	7,8
en	10	5,3	10	7,3	7	5,5
N <sub>(c)</sub> N C N <sub>c</sub>	3	1,6	3	2,2	3	2,3
Yhteensä	180	95,3	131	95,6	123	96,1

Huom. Verbien partisiippi muodot substantiivien määritteinä on luokiteltu adjektiiveiksi.

Ydintermien eli kaikkein varminten termien tunnistamiseksi vertasin asiantuntijoiden ensimmäisellä kierroksella yksimielisesti valitsemien 118 termin listaa opiskelijoiden termiehdokaslistoihin ja katsoin, ovatko ne myös opiskelijoiden mielestä varminten termit. Oletettavasti termit, jotka asiantuntijat valitsivat heti ensimmäisellä kierroksella ja jotka myös noin kaksi kolmasosaa koehenkilöistä on valinnut, ovat tekstin ydintermit. Merenkulun opiskelijoiden listassa näitä vähintään seitsemän opiskelijan valitsemia termit on 101 ja kääntämisen opiskelijoiden listassa 132.

Vähintään seitsemän merenkulun opiskelijan ja asiantuntijoiden heti ensimmäisellä kierroksella valitsemista termeistä kiinnittää huomiota niiden termien osuus, joissa termi tai terminosa on lyhenne. 66 termistä 17 noudattaa termikaavaa Abb (*VTS*) tai Abb N (*VHF radio*). Vähintään seitsemän kääntämisen opiskelijan ja asiantuntijoiden ensimmäisellä kierroksella valitsemien termien listassa on 82 termiä, joista 28 termistä on lyhenne. Lyhenteet siis ovat selkeästi termit. Termikaavan N<sub>c</sub> N C N<sub>c</sub> mukaisissa termeissä (*laivaliikenteen ohjaus- ja informaatiojärjestelmä, laivaliikenteen ohjaus- ja informaatiopalvelu, meriliikenteen ohjaus- ja informaatiopalvelu*) on jotakin erikoista, koska kaikki kolme tämän termikaavan mukaista termiä on poimittu lähes yksimielisesti. Nämä ovat esitelmän aiheen kannalta keskeisiä termit ja itse asiassa keskeisen lyhenteen avattu muotoja.

Keskimääräiseltä sanaluvultaan, frekvenssiltään tai termikaavoiltaan kääntämisen opiskelijoiden ja asiantuntijoiden ensimmäisellä kierroksella valitsemat termit eivät juuri eroa merenkulun opiskelijoiden ja asiantuntijoiden ensimmäisellä kierroksella valitsemista termeistä. Erojen syitä onkin etsittävä termivariaatiosta ja käsitetasolta. Seuraavaksi etsin vastausta kysymykseen, millaisia ovat pituudeltaan, frekvenssiltään ja kie-

liopilliselta rakenteeltaan ne termit, jotka asiantuntijat poimivat mutta opiskelijat jättivät termiehdokaslistan ulkopuolelle.

#### 6.4.2 Suomalaisen opiskelijaryhmien piilotermien ominaisuuksia

Suomalaiset kääntämisen ja merenkulun opiskelijat poimivat termiehdokaslistoihinsa suurimman osan suomalaisten asiantuntijoiden valitsemista termeistä. Asiantuntijalistalla 189 termistä kääntämisen opiskelijoiden listassa on 137 ja merenkulun opiskelijoiden listassa 128. Opiskelijoiden listoista kuitenkin puuttuu noin kolmannes termeistä: merenkulun opiskelijoiden termiehdokaslistasta 61 ja kääntämisen opiskelijoiden listasta 52 asiantuntijoiden valitsemia termiä. Tässä jaksossa tarkastelen suomalaisten opiskelijoiden termiehdokaslistoista puuttuvien termien sanalukua, frekvenssiä ja termikaavoja. Tavoitteena on vastata kysymykseen, onko koehenkilöryhmien piilotermieissä sellaisia eroja, jotka selittyvät koehenkilöiden erilaisella koulutustaustalla.

Sanaluvultaan merenkulkijoiden termiehdokaslistasta puuttuvat termit ovat keskimäärin 1,62 sanan mittaisia. Suurin osa puuttuvista termeistä on yhden sanan mittaisia, kuten suurin osa termeistäkin. Vähintään kolmen sanan pituisia termejä puuttuu kuusi (*etsintä ja pelastus*, *Long Range Tracking*, *maalle rakennettava tutka-asema*, *maalle rakennettu tutka-asema*, *Maritime Industries Forum*, *Vessel Traffic Management and Information System*). Kääntämisen opiskelijoiden termiehdokaslistasta puuttuvat termit ovat sanaluvultaan keskimäärin 1,35 sanan mittaisia. Vähintään kolmen sanan pituisia termejä puuttuu vain kaksi (*etsintä ja pelastus*, *Maritime Industries Forum*).

TAULUKKO 13. Suomalaisen koehenkilöryhmien piilotermien pituus

Sanaluku	kofi (A=1,35)		mofi (A=1,62)	
	f	%	f	%
1	36	69,2	32	52,5
2	14	26,9	23	37,7
3	2	3,9	5	8,2
≥ 4	0	0,0	1	1,6
Yhteensä	52	100	61	100

Kummassakin ryhmässä suurin osa piilotermieistä on yksisanaista: kääntämisen opiskelijoiden ryhmässä liki 70 prosenttia. Merenkulun opiskelijoiden termiehdokaslistasta puuttuu yksisanaisten lisäksi runsaasti kaksisanaista termejä, sillä niiden osuus puuttuvista termeistä on liki 38 prosenttia.

Kääntämisen opiskelijoiden termiehdokaslistasta puuttuvat termit esiintyvät lähtötekstissä keskimäärin 1,56 kertaa, kun asiantuntijoiden kanssa samoin poimituilla termeillä on keskimäärin 2,58 esiintymää. Suurin osa kääntämisen opiskelijoiden termiehdokaslistasta puuttuvista termeistä onkin yhden esiintymän termejä. Keskimääräistä esiintymien määrää kasvattaa muutama termiehdokaslistasta puuttuva suurifrekvenssinen termi (*liikenne*, *ohjaus*, *onnettomuus*, *tehokkuus*, *tietojen vaihto*, *turvallisuus*, *VTS menetelmä*), jotka esiintyvät lähtötekstissä vähintään viisi kertaa. Piilotermien joukossa on myös peräti 15 nollafrekvenssistä termiä eli termiä, joka ei esiinny tekstissä yhtään kertaa siinä muodossa, jossa se on termilistassa, vaan joko toisen termin osana tai elliptisenä. Kääntämisen opiskelijoilla oli siis eniten vaikeuksia lyhyiden ja pienifrekvenssisten termien poiminnassa. Ryhmältä jäivät poimimatta kaikki nollafrekvenssiset termit, esimerkiksi termit *etsintä ja pelastus*, *hinaus*, *lastaus* ja *luotsipalvelu*.

Merenkulun opiskelijoiden termiehdokaslistan piilotermit esiintyvät tekstissä keskimäärin vain yhden kerran mutta asiantuntijoiden kanssa samoin valituilla termeillä on keskimäärin lähes kolme esiintymää. Suuri osa piilotermeistä on yhden esiintymän termejä. Myös merenkulun opiskelijoiden termiehdokaslistasta puuttuu 15 nollafrekvenssistä termiä. Vähintään kolmen esiintymän termejä puuttuu vain kaksi (*tietojen vaihto, yhteistoimintakomitea*). Merenkulun opiskelijat poimivat suurifrekvenssiset termit tarkemmin kuin kääntämisen opiskelijat, jotka taas poimivat tarkemmin pienifrekvenssiset termit. Merenkulun opiskelijoilla oli vaikeuksia lyhyiden, pienifrekvenssisten termien poiminnassa. Varsinkin nollafrekvenssiset termit tuottivat hankaluuksia.

TAULUKKO 14. Suomalaisten koehenkilöryhmien piilotermin frekvenssi

Frekvenssi	kofi (A=1,56)		mofi (A=1,05)	
	f	%	f	%
$\leq 1$	38	73,0	47	77,1
2	7	13,5	12	19,7
3	0	0,0	1	1,6
$\geq 4$	7	13,5	1	1,6
Yhteensä	52	100	61	100

Merenkulkijoiden termiehdokaslistasta puuttuu eniten substantiivitermejä, joiden termikaavat ovat  $N_c$ ,  $N$  tai  $N_{(c)}$   $N_{(c)}$ . Myös adjektiivin ja substantiivin muodostamista sanaliittotermeistä puuttuu kaksi kolmasosaa. Muiden tavallisempien termikaavojen Abb-N, Abb N, en ja Abb mukaisista termeistä puuttuu vain muutama. Lyhenteistä puuttuu vain yksi, *MIF*, ja termikaavaltaan ”kompleksisista” termeistä eli vähintään kolmen sanan mittaisista termeistä puuttuu vain kolme suomenkielistä termiä.

Kääntämisen opiskelijoiden termiehdokaslistasta puuttuvat termit noudattavat merenkulun opiskelijoiden listasta puuttuvien termien kaavoja. Kääntämisen opiskelijoiden listasta puuttuu eniten yhden kantasanan substantiiveja  $N$  ja toiseksi eniten yhdyssubstantiiveja  $N_c$ . Kahden substantiivin sanaliittotermeistä  $N_{(c)}$   $N_{(c)}$  puuttuu lähes joka toinen. Adjektiivin ja substantiivin sanaliittotermin  $A N_{(c)}$  sekä lyhenteen ja substantiivin muodostamien termien Abb N tai Abb-N poiminnassa kääntämisen opiskelijoiden poiminta on lähempänä asiantuntijoiden poimintaa kuin merenkulun opiskelijoiden, sillä näiden termikaavojen mukaisista termeistä poimimatta jäi vain kolme. Englanninkielisistä termeistä kääntämisen opiskelijoilta jäi yksi poimimatta. Lyhenteistä puuttuu vain lyhenne *MIF* ja vähintään kolmen sanan mittaisista termeistä yksi, *etsintä ja pelastus*. Kääntämisen opiskelijat eivät kuitenkaan poimineet asiantuntijalistan ainoaa verbitermiä *ohjailla*.

TAULUKKO 15. Suomalaisen koehenkilöryhmien piilotermin tavallisimmat termikaavat

Termikaava	kofi		mofi	
	f	%	f	%
N <sub>c</sub>	16	30,8	16	26,2
Abb N, Abb-N	1	1,9	4	6,6
N	18	34,6	12	19,7
N <sub>(c)</sub> N <sub>(c)</sub>	10	19,2	10	16,4
A N <sub>(c)</sub>	3	5,8	11	18,0
Yhteensä	49	94,2	57	93,5

Kummankin koehenkilöryhmän termiehdokaslistasta puuttuu eniten lyhyitä termejä, jotka noudattavat yleisimpiä suomalaisten termien kaavoja N, N<sub>c</sub> ja N<sub>(c)</sub> N<sub>(c)</sub>. Näiden kaavojen mukaisia piilotermejä puuttuu kummankin ryhmän termiehdokaslistasta lähes yhtä monta. Ryhmien välinen ero näkyy selvimmin adjektiivin ja substantiivin muodostamien sanaliittotermin poiminnassa. Tämän kaavan mukaisista termeistä puuttuu merenkulun opiskelijoiden listasta noin kaksi kolmasosaa, mutta kääntämisen opiskelijoiden listasta vain noin yksi viidesosa. Merenkulun opiskelijoilla oli siis vaikeuksia poimia adjektiivin ja substantiivin muodostamia sanaliittotermejä, joiden poiminta onnistui kääntämisen opiskelijoilta varsin hyvin. Lyhenteen sisältävien termien poiminta onnistui kummaltakin ryhmältä hyvin.

Kääntämisen opiskelijoiden termiehdokaslistasta puuttuvista termeistä noin puolet eli 30 on samoja, jotka myös merenkulun opiskelijat ovat jättäneet poimimatta (*etsintäviranomais*, *kuljetusketju*, *kulku*, *laivan käsittely*, *laivojen jäljittäminen*, *laivojen tunnistaminen*, *lastaus*, *logistinen palvelu*, *luotsipalvelu*, *merenkulku*, *pelastusviranomais*, *päällystä*, *radio*, *rannikkovartiosto*, *sensori*, *tietojen kulku*, *tietojen vaihto*). Merenkulun opiskelijoista poiketen kääntämisen opiskelijat eivät pitäneet termeinä esimerkiksi termejä *etsintä*, *laivan sijainti*, *lastin käsittely*, *näkyvyys*, *ohjailla*, *ohjaus*, *onnettomuus*, *pelastus*, *sijainti*, *vaarallinen aine* ja *vesikuljetus*. Näiden termien poimintaan siis merenkulualan koulutus ilmeisesti on vaikuttanut.

Merenkulun opiskelijoilta poimimatta jääneissä termeissä on usein yhteinen termin osa jonkin opiskelijoiden poimiman termiehdokkaan kanssa. Merenkulun opiskelijat eivät poimineet esimerkiksi termiä *etsintä* ja *pelastus*, mutta kylläkin sen osat *etsintä* ja *pelastus*, jotka myös ovat termejä, eivätkä termejä *laivakäyttöinen tutka* tai *tutka-aseman operaattori*, mutta he poimivat termit *tutka* ja *operaattori*. Merenkulun opiskelijoiden mielestä määriteosat *laivakäyttöinen* ja *tutka-asema* eivät siis ole termin osia. Merenkulun opiskelijoiden termiehdokaslistasta puuttuu englanninkielisiä termejä (*Allied Services*, *Long Range Tracking*), termejä, jotka ovat myös yleiskielen sanoja (*asema*, *hinaus*, *jakelu*, *kulku*, *lastaus*, *ohjeistus*, *merenkulku*, *radio*, *sensori*, *varastointi*), ja muiden alojen kuin merenkulun termejä (*elektroninen tietojenvaihto*, *raportointimuoto*, *sataman käyttöaste*, *strateginen liikennekuva*, *tietojen kulku*, *tietojen vaihto*, *tietokanta*, *vertikaalinen integraatio*, *vertikaalinen tietojenvaihto*, *yhdistyneet palvelut*, *yhteistyötoimikunta*).

Merkillepantavaa on, että kummankin suomalaisen ryhmän koehenkilöt poimivat englanninkielisiä termejä, englanninkielisiä lyhenteitä sekä englanninkielisten lyhenteiden ja suomenkielisten substantiivien yhdistelmiä varsin samalla tavalla kuin suomalaiset asiantuntijat. Kokonaan tai osittain englanninkielisten termien saanti on kummassakin koehenkilöryhmässä varsin korkea: kääntämisen opiskelijoiden ryhmässä se on 93,5 ja merenkulun opiskelijoiden ryhmässä 80,4 prosenttia. Kääntämisen opiskelijoiden



listasta puuttuu vain yksi englanninkielinen termi, *Maritime Industries Forum*, yksi englanninkielinen lyhenne, *MIF*, ja yksi englanninkielisen lyhenteen ja suomenkielisen substantiivin yhdistelmä, *VTS menetelmä*. Merenkulun opiskelijoiden listasta puuttuu sama lyhenne kuin kääntämisen opiskelijoiden listasta. Englanninkielisiä termejä sekä lyhenteen ja substantiivin yhdistelmiä puuttuu kumpiakin neljä.

#### 6.4.3 Suomalaisten opiskelijaryhmien poimimien valetermien ominaisuuksia

Seuraavaksi tarkastelen niitä kustakin opiskelijaryhmästä vähintään neljän koehenkilön valitsemia termiehdokkaita, joita asiantuntijat eivät valinneet ja joita siis nimitän valetermeiksi. Tarkasteltavia piirteitä ovat valetermien pituus sanaluvulla mitattuna, frekvenssi ja termikaavat. Tavoitteena on vastata kysymyksiin, millaisia ovat toisaalta kääntämisen opiskelijoiden ja toisaalta merenkulun opiskelijoiden valitsevat valetermit ja miten koehenkilöryhmien valitsevat valetermit eroavat sanaluvultaan, frekvenssiltään tai termikaavoiltaan.

Merenkulun opiskelijoiden termiehdokaslistassa on 63 valetermiä, joiden pituus on keskimäärin 1,7 sanaa. Suurin osa niistä on yhden sanan mittaisia. Vähintään kolmen sanan mittaisia valetermejä on 10 kappaletta. Nämä joko ovat kahden termin yhdistelmiä tai sisältävät termiin kuulumattomia määritteitä. Esimerkiksi termiehdokkaassa *meri- ja vesikuljetus* on kaksi termiä, *merikuljetus* ja *vesikuljetus*, ja termiehdokkaassa *vii-veetön ja häiriötön liikenne* on määriteosia, jotka eivät kuulu termiin, sillä asiantuntijoiden mukaan termiin kuuluu vain pääsana *liikenne*.

Kääntämisen opiskelijoiden valitsemien valetermien sanaluku ei juuri eroa merenkulun opiskelijoiden valitsemien valetermien sanaluvusta, sillä kääntämisen opiskelijoiden valitsevat valetermit ovat keskimäärin 1,83 sanan mittaisia. Niistä noin puolet on yhden sanan mittaisia. Kääntämisen opiskelijat poimivat kaikenpituisia valetermejä lukumääräisesti enemmän kuin merenkulun opiskelijat. Yksisanaisten ohella he poimivat runsaasti kaksisanaisia valetermejä ja vähintään kolmisanaisia valetermejäkin liki kaksi kertaa niin paljon kuin merenkulun opiskelijat. Nämä valetermit ovat termien yhdistelmiä tai niissä on termiin kuulumattomia osia aivan samoin kuin merenkulun opiskelijoiden valetermeissä.

TAULUKKO 16. Suomalaisten koehenkilöryhmien poimimien valetermien pituus

Sanaluku	kofi (A=1,83)		mofi (A=1,71)	
	f	%	f	%
1	48	47,5	39	61,9
2	34	33,7	14	22,3
3	13	12,9	5	7,9
≥ 4	6	5,9	5	7,9
Yhteensä	101	100	63	100

Kääntämisen opiskelijoiden poimimat valetermit esiintyvät tekstissä keskimäärin 1,39 kertaa. Suurin osa valetermeistä eli 76 prosenttia esiintyy tekstissä vain yhden kerran. Lisäksi kääntämisen opiskelijat poimivat neljä nolلافrekvenssistä termiehdokasta, jollainen on esimerkiksi *merenkulun toimija*, kolme sellaista yleiskielistä substantiivia, jotka esiintyvät tekstissä vähintään viisi kertaa, kuten esimerkiksi substantiivit *menetelmä* ja *säännös*, sekä erisnimen lyhenteen *EU*.

Merenkulun opiskelijoiden poimimat valetermit esiintyvät tekstissä keskimäärin 2,13 kertaa. Valetermien keskimääräinen frekvenssi on korkeampi kuin kääntämisen opiskelijoilla, koska merenkulun opiskelijat poimivat huomattavasti vähemmän yksifrekvenssisiä vatermejä. Suurin osa vatermeistä on kuitenkin yksifrekvenssisiä kuten kääntämisen opiskelijoillakin. Vähintään viisi kertaa tekstissä esiintyviä vatermejä merenkulun opiskelijoiden listassa on kahdeksan. Niistä viisi on yleiskielisiä substantiiveja ja kolme adjektiiveja, joita asiantuntijoiden termilistassa ei ole lainkaan. Merenkulun opiskelijoiden termiehdokaslistassa on vain yksi sellainen termiehdokas, jolla on tekstissä nolafrekvenssi. Termiehdokas *meriliikenteen ohjaus* esiintyy tekstissä vain osana termiä *meriliikenteen ohjausjärjestelmä* ja sanaliittoa *meriliikenteen ohjauspyrkimys*.

TAULUKKO 17. Suomalaisten opiskelijaryhmien poimimien vatermien frekvenssi

Frekvenssi	kofi (A=1,39)		mofi (A=2,13)	
	f	%	f	%
≤ 1	81	80,2	43	68,2
2	11	10,9	10	15,9
3	3	3,0	1	1,6
≥ 4	6	5,9	9	14,3
Yhteensä	101	100	63	100

Toisin kuin sanaluvussa ja frekvenssissä termikaavoissa on koehenkilöryhmien välillä huomattavia eroja. Merenkulun opiskelijoiden poimimien vatermien joukossa on eniten yhdyssubstantiiveja  $N_c$ , kahden substantiivin sanaliittoja  $N_{(c)}$   $N_{(c)}$  ja substantiiveja  $N$ . Adjektiivitermiehdokkaita on huomattavan paljon mutta adjektiivin ja substantiivin muodostamia sanaliittoja on vain yksi. Lyhenteen ja substantiivin muodostamia termiehdokkaita ja englanninkielisiä sanaliittoja on vatermien listassa kumpiakin viisi. Näiden termikaavojen mukaisten termien valinta ei ole osunut yksiin asiantuntijoiden valinnan kanssa.

Silmiinpistävin piirre kääntämisen opiskelijoiden valinnassa on sellaisten vatermien suuri määrä, joissa termin osana on lyhenne. Tällaisia vatermejä listassa on 31 kappaletta. Huomattava on myös adjektiivin ja substantiivin tai yhdyssubstantiivin muodostamien sanaliittojen suuri osuus.

TAULUKKO 18. Suomalaisten opiskelijaryhmien poimimien vatermien tavallisimmat termikaavat

Termikaava	kofi		mofi	
	f	%	f	%
$N_c$	27	26,7	11	17,5
Abb X	31	30,7	5	7,9
N	7	6,9	9	14,3
$N_{(c)}$ $N_{(c)}$	11	10,9	11	17,5
A $N_{(c)}$	12	11,9	1	1,6
X C X	8	7,9	5	7,9
Yhteensä	96	95,5	42	66,7

Yhteenvedon merenkulun opiskelijoiden vaalestermistä voidaan todeta, että ne voivat olla liian lyhyitä, kuten adjektiivit *alueellinen*, *integroitu* ja *kansainvälinen*, tai liian pitkiä termien yhdistelmiä, kuten *etsintä-*, *pelastus-* ja *ympäristönsuojeluviranomaiset*, jossa on kolme termiä, *etsintäviranomaisen*, *pelastusviranomaisen* ja *ympäristönsuojeluviranomaisen*. Sanaliitossa *liikenteenohjausjärjestelmä* (VTS) on kaksi termiä, *liikenteenohjausjärjestelmä* ja VTS. Viiden sanan termiehdokas *luotsien ja tutka-asemien operaattoreiden yhteistoiminta* jakautuu neljäksi termiksi, *luotsi*, *tutka-asema*, *operaattori* ja *tutka-aseman operaattori*. Liian pitkät termiehdokkaat voivat myös sisältää termiin kuulumattomia määritteitä. Esimerkiksi termiehdokkaassa *rantavaltion vilkasliikenteinen aluemi* on määriteosia, jotka eivät kuulu termiin, sillä asiantuntijoiden mukaan termiin kuuluu vain pääsana *aluemi*. Merenkulun opiskelijat valitsivat termiehdokkaiksi myös yksilökäsitteiden nimityksiä: *Elbe-joki*, *Englannin kanaali* ja *Hampurin satama*.

Terminpoiminnan yksimielisyydeksi merenkulun opiskelijoiden termiehdokaslistassa saadaan 5,27 (332 jaettuna 63:lla), eli keskimäärin vain noin viisi opiskelijaa 11:sta valitsi termiehdokkaat, joita asiantuntijat eivät valinneet. Asiantuntijoiden kannasta poiketen moni merenkulun opiskelija valitsi termiehdokkaat *merenkulkumaa* (10), *merenkulkutausta* (8), *merikokemus* (9), *vilkasliikenteinen* (7) ja *ympäristönsuojelu* (7). Su-luissa on nämä termiehdokkaat valinneiden opiskelijoiden määrä.

Kääntämisen opiskelijoiden vaalestermistä moni oli liian lyhyt ja sellaisena laajemman käsitteen nimitys kuin asiantuntijoiden valitsema termi (*käsikirja*, *käyttöaste*, *laivakäyttöinen*, *liikennekuva*, *menetelmä*, *off-shore*, *VHF*), mutta tätäkin useammin kääntämisen opiskelijoiden poimimat vaalestermit ovat kahden tai kolmen termin yhdistelmiä. Esimerkiksi sanaliitto *laivojen ilmaiseminen, tunnistaminen ja jäljittäminen* on yhdistelmä kolmesta termistä, *laivojen ilmaiseminen*, *laivojen tunnistaminen* ja *laivojen jäljittäminen*. Sanaliitoissa *rantavaltion aluemi* ja *turvallisuus- ja ympäristöriski* on kummassakin kaksi termiä: *rantavaltio* ja *aluemi* sekä *turvallisuusriski* ja *ympäristöriski*. Kääntämisen opiskelijoiden termiehdokaslistassa on myös merenkulun opiskelijoiden kanssa samoin valittuja liian pitkiä termiehdokkaita, kuten *liikenteenohjausjärjestelmä* (VTS) ja *luotsien ja tutka-asemien operaattoreiden yhteistoiminta*, joihin sisältyy useampi kuin yksi termi. Myös kääntämisen opiskelijoiden listassa on yksilökäsitteiden nimityksiä tai nimikkeitä, kuten *4:s puiteohjelma*, *APAS-silta*, *COST 301-tutkimus*, *EU*, *EURET* ja *IALA:n johtokunta*, lyhenteen ja substantiivin yhdistelmiä, kuten *super-VTS*, *VTMIS-käsite*, *VTS määrite*, tai liian pitkiä termiehdokkaita. Sanaliitoissa *EU:n HAZMAT-direktiivi*, *kansallinen karttakorjauspalvelu*, *määräaikainen VTS komitea* ja *nykyaikainen VTS toiminta* termiehdokkaan alussa on yksi ylimääräinen määrite, kun taas sanaliitoissa *VTS menetelmien ja säännösten kehittäminen*, *merikuljetusten tehostaminen* ja *meriliikenteen ohjauspyrkimys* ylimääräistä onkin pääsana. Termejä ovat asiantuntijoiden mukaan *VTS menetelmä*, *VTS säännös*, *merikuljetus* ja *meriliikenne*.

Lisäksi kääntämisen opiskelijoiden terminpoiminnan yksimielisyydeksi merenkulun termienä, vaikka ne ovatkin kielellisesti mahdollisia termejä. Tällaisia termiehdokkaita ovat esimerkiksi *funktionaalinen määritelmä*, *määrite*, *perustoiminto*, *päätoimija*, *toimintamääritelmä* ja *viitetermi*. Ehkä yhdyssanojen ja sanaliittojen osina esiintyvät sanat *määritelmä* ja *termi* ovat kiinnittäneet opiskelijoiden huomion, koska nämä sanat ovat termejä esimerkiksi kielitieteessä. Sanojen kontekstissa on myös saattanut olla jotakin sellaista, minkä vuoksi ne on valittu termiehdokkaiksi.

Kääntämisen opiskelijoiden terminpoiminnan yksimielisyydeksi vaalestermien listassa saadaan 5,56 (562 jaettuna 101:lla) eli keskimäärin viisi tai kuusi opiskelijaa 11:sta va-

litsi termiehdokkaat, joita asiantuntijat eivät valinneet. Asiantuntijoiden kannasta poiketen moni kääntämisen opiskelija valitsi termiehdokkaat *COST 301-tutkimus*, *EURET*, *Just On and In Time periaate*, *kuljetuslogistiikkaverkosto*, *operaattorikoulutus*, *VTMIS-käsite* ja *VTMIS-tutkimus*, joita 10 kääntämisen opiskelijaa piti termeinä.

Edellä kuvattujen piirteiden perusteella koehenkilöiden valitsemat valetermit voidaan jakaa neljään ryhmään:

- 1) termiehdokkaat, joissa on ylimääräinen määräiteosa, kuten esimerkiksi *uusi järjestelmä*,
- 2) rakenteeltaan termeiksi sopivat mutta liian pitkät sanaliitot, kuten esimerkiksi *tutkaverkko sisääntuloväylän varrella*, *Malaccan salmen liikenteen valvominen*,
- 3) rakenteeltaan termeiksi sopimattomat sanaliitot, kuten esimerkiksi *muu alalla toimiva*, *ylös Elbe-jokea kohti Hampurin satamaa*,
- 4) yleiskielen sanat tai sanaliitot, jotka eivät ole merenkulun termejä, kuten esimerkiksi *tulla*, *varten*, *kansainvälinen määräite*, *kansallinen taso*.

## 6.5 Venäläisten opiskelijaryhmien terminpoiminta

Jaksossa 6.4 selostin suomalaisten opiskelijaryhmien terminpoimintaa ja vertasin koehenkilöryhmien valintaa asiantuntijoiden valintaan. Tavoitteena oli selvittää, miten koehenkilöiden erilainen koulutustausta vaikuttaa terminpoimintaan. Seuraavissa jaksoissa selostan vastaavaa vertailua venäläisten opiskelijaryhmien ja venäläisten asiantuntijoiden terminpoiminnan välillä. Tavoitteena on selvittää, ovatko erot samansuuntaisia kuin suomalaisten koehenkilöryhmien väliset erot.

Venäläisten opiskelijaryhmien välistä vertailua varten kokosin kummankin ryhmän termiehdokaslistoista yhteislistan, jossa ovat vähintään neljän koehenkilön valitsemat termiehdokkaat. Tämän kriteerin täyttää 161 venäläisten kieliaineiden opiskelijoiden valitsemaa termiehdokasta ja 253 merenkulun opiskelijoiden valitsemaa termiehdokasta. Vähintään neljän koehenkilön valitsemia termiehdokkaita on kummankin koehenkilöryhmän termiehdokaslistoissa vain noin neljännes. Kieliaineiden opiskelijoiden ryhmässä vähintään neljän koehenkilön valitsemien termiehdokkaiden osuus on 28,4 prosenttia ja merenkulun opiskelijoiden ryhmässä 24,3 prosenttia. Yhteislistoista muodostin ryhmittäin suppeammat termiehdokaslistat, joissa ovat opiskelijoiden valitsemat termit, opiskelijoiden listoista puuttuvat piilotermit ja opiskelijoiden valitsemat valetermit.

Seuraavissa jaksoissa selostan, millaisia ovat pituudeltaan, frekvenssiltään ja termikaavoiltaan sekä asiantuntijoiden että opiskelijoiden valitsemat termit, opiskelijaryhmien termiehdokkaiden joukosta puuttuvat piilotermit ja opiskelijaryhmien poimimat valetermit. Venäläisten kieliaineiden opiskelijoiden ryhmästä käytän lyhennettä *koru* ja merenkulun opiskelijoiden ryhmästä lyhennettä *moru*.

### 6.5.1 Venäläisten opiskelijaryhmien poimimien termien ominaisuuksia

Tarkastelen seuraavaksi niitä venäläisten asiantuntijoiden valitsemia termejä, jotka myös vähintään neljä venäläistä kieliaineiden opiskelijaa tai vähintään neljä venäläistä merenkulun opiskelijaa valitsi termiehdokkaaksi. Tässäkin tarkastelussa kiinnitän huomiota termien pituuteen sanaluvulla mitattuna, frekvenssiin ja termikaavoihin.

Vähintään neljän venäläisen kieliaineiden opiskelijan valitsemien termien keskimääräinen sanaluku on 2,07 eli hieman alhaisempi kuin venäläisten asiantuntijoiden valitsemien termien keskimääräinen sanaluku, joka on 2,36. Merenkulun opiskelijoiden va-

litsemien termien keskimääräinen sanaluku on 1,96. Erot koehenkilöryhmien välillä ovat pienet, sillä kummankin ryhmän koehenkilöt valitsivat lähes saman verran eripituisia termejä. Sen sijaan ero asiantuntijoiden valintaan on selkeä. Venäläiset koehenkilöryhmät valitsivat liki kaikki asiantuntijoiden valitsemat yksisanaiset termit, mutta saanti on sitä pienempi mitä pitempi termi on. Vähintään nelisanaisten termien saanti kieliaineiden opiskelijoiden ryhmässä on 37 prosenttia mutta merenkulun opiskelijoiden ryhmässä vain 25 prosenttia.

TAULUKKO 19. Venäläisten asiantuntijoiden ja opiskelijaryhmien poimimien termien pituus

Sanaluku	atru (A=2,36)		koru (A=2,07)		moru (A=1,96)	
	f	%	f	%	f	%
1	28	26,1	26	31,7	27	36,5
2	40	37,4	35	42,7	30	40,5
3	23	21,5	15	18,3	13	17,6
≥ 4	16	15,0	6	7,3	4	5,4
Yhteensä	107	100	82	100	74	100

Venäläisten koehenkilöiden valitsemien termien keskimääräisessä frekvenssissä ryhmien välillä on vain pieni ero mutta asiantuntijoiden valitsemien termien frekvenssiin verrattuna erot ovat selkeät. Samoin kuin suomalaisten myös venäläisten koehenkilöiden valitsemien termien keskimääräinen frekvenssi on hieman korkeampi kuin asiantuntijoiden valitsemien. Kieliaineiden opiskelijoiden ryhmässä keskimääräinen frekvenssi on 4,57 ja merenkulun opiskelijoiden ryhmässä 4,65, kun se asiantuntijoiden valinnassa on 3,95. Ero selittyy sillä, että koehenkilöryhmiltä jäi poimimatta suhteessa eniten pienifrekvenssisia ( $f \leq 1$ ) termejä. Kieliaineiden opiskelijat onnistuivat näiden termien poiminnassa hieman paremmin kuin merenkulun opiskelijat. Sen sijaan suurifrekvenssisten termien ( $f \geq 4$ ) poiminnassa koehenkilöt ovat onnistuneet erityisen hyvin. Tämä tulos poikkeaa suomalaisten koehenkilöiden valinnasta, sillä suomalaisten koehenkilöiden valitsemisissa termeissä suurifrekvenssisten termien osuus on huomattavasti pienempi kuin suomalaisten asiantuntijoiden valinnassa.

TAULUKKO 20. Venäläisten asiantuntijoiden ja opiskelijaryhmien poimimien termien frekvenssi

Frekvenssi	atru (A=3,95)		koru (A=4,57)		moru (A=4,65)	
	f	%	f	%	f	%
≤ 1	56	52,3	39	47,6	34	45,9
2	14	13,1	11	13,4	11	14,9
3	9	8,4	8	9,7	7	9,5
≥ 4	28	26,2	24	29,3	22	29,7
Yhteensä	107	100	82	100	74	100

Termikaavaltaan koehenkilöryhmien valitsemat termit eivät juuri eroa toisistaan eivätkä myöskään venäläisten asiantuntijoiden valinnasta. Yleisin termikaava on adjektiivin ja substantiivin sanaliitto A N, joita termeistä on kummassakin ryhmässä noin neljännes.

Koehenkilöryhmiltä jäi poimimatta kaksi tai kolme tämän termikaavan mukaisista termeistä. Toiseksi eniten on yksisanaisia substantiivitermejä, joista kieliaineiden opiskelijoilta jäi poimimatta yksi (*бедствие*). Kolmanneksi eniten kummankin koehenkilöryhmän valitsemista termeistä on lyhenteitä. Kummankin ryhmän koehenkilöt poimivat termiehdokaslistoihinsa 11 asiantuntijoiden valitsemista 12 lyhenteestä. Vain lyhenne *GPS* puuttuu termiehdokaslistoista. Toisin kuin suomenkielisestä tekstistä valituista termeistä venäjänkielisestä tekstistä valituista termeistä vain kolme lyhennetermiä on englanninkielistä. Ne ovat lyhenteet *DGPS*, *GPS* ja *VTMIS*, jotka myös koehenkilöt valitsivat termiehdokkaiksi *GPS*-lyhennettä lukuun ottamatta. Yllättäen kummallakin ryhmällä oli vaikeuksia myös kahdesta substantiivista koostuvien sanaliittotermien poiminnassa. Esimerkiksi termi *зосударство флага* on kummankin koehenkilöryhmän piilotermien listalla. Selkein ero ryhmien välillä on kahden adjektiivin ja yhden substantiivin muodostamien sanaliittojen valinnassa, sillä tämän termikaavan mukaisista seitsemästä termistä kieliaineiden opiskelijat poimivat kuusi mutta merenkulun opiskelijat vain kolme. Esimerkiksi *морская спутниковая связь* puuttuu kummankin ryhmän termitlistasta.

TAULUKKO 21. Venäläisten asiantuntijoiden ja opiskelijaryhmien poimimien termien tavallisimmat termikaavat

Termikaava	atru		koru		moru	
	f	%	f	%	f	%
A N	22	20,6	20	24,4	19	25,7
N	16	15,0	15	18,3	16	21,6
Abb	12	11,2	11	13,4	11	14,8
N N	13	12,1	9	11,0	10	13,5
A A N	7	6,5	6	7,3	3	4,1
Yhteensä	74	69,1	65	79,3	62	83,8

Venäläiset asiantuntijat valitsivat yksimielisesti heti ensimmäisellä kierroksella 37 termiä. Näistä 36 on venäläisten asiantuntijoiden ja venäläisten koehenkilöiden yksimielisesti valitsemien termien joukossa. Yksi asiantuntijoiden heti ensimmäisellä kierroksella valitsema termi puuttuu merenkulun opiskelijoiden valitsemien termien listasta. Tämä termi on *несение вахты* (vahdinpito). Merenkulkijat ovat ehkä pitäneet päivittäiseen työhön kuuluvaa toimintaa niin arkisena, etteivät ole pitäneet sitä terminä. Asiantuntijoiden heti ensimmäisellä kierroksella valitsemia termejä voidaan siis hyvällä syyllä pitää ydintermeinä. Samoin kuin suomalaisten ydintermien listassa tässäkin joukossa on suurifrekvenssisia lyhenteitä, kuten *АИС* (AIS) ja *СУДС* (VTS), sekä yleiskielen sanoja, kuten *порт* (satama), *судно* (alus) ja *судоходство* (alusliikenne), mutta myös vain kerran lähtötekstissä esiintyviä termejä, kuten *открытое море* (avomeri), *подходный путь* (sisääntuloväylä) ja *судозаход* (satamakäynti).

### 6.5.2 Venäläisten opiskelijaryhmien piilotermien ominaisuuksia

Venäläisten koehenkilöryhmien terminpoiminnan saanti on pari prosenttiyksikköä korkeampi kuin vastaavan suomalaisen koehenkilöryhmän. Kummassakin kielessä merenkulun opiskelijoiden termiehdokaslistoista puuttuu hieman enemmän termejä. Venäläisten merenkulun opiskelijoiden poiminnassa piilotermiä on 33 ja kieliaineiden opiskelijoiden poiminnassa 25. Piilotermien pituudessa ei koehenkilöryhmien välillä ole suurta



eroa. Merenkulun opiskelijoiden piilotermit ovat keskimäärin 3,24 ja kieliaineiden opiskelijoiden keskimäärin 3,28 sanan mittaisia. Yksisanaisia termejä puuttuu kummankin ryhmän termiehdokaslistasta vain vähän: merenkulun opiskelijoiden listasta puuttuu lyhenne-termi *GPS* ja kieliaineiden opiskelijoiden listasta substantiivi *бедствие* (häätä) ja lyhenne *GPS*. Selkein ero ryhmien välillä on kaksisanaisten termien poiminnassa, jossa kieliaineiden opiskelijat onnistuivat merenkulun opiskelijoita paremmin. Piilotermeistä kaksi kolmasosaa on vähintään kolmisanaisia kummankin ryhmän valinnassa. Merenkulun opiskelijoiden listasta pitkiä termejä puuttuu 22 ja kieliaineiden opiskelijoiden listasta 18. Pituutta kasvattavat prepositiot, jotka on tässä tutkimuksessa laskettu sanoiksi.

TAULUKKO 22. Venäläisten koehenkilöryhmien piilotermin pituus

Sanaluku	koru (A=3,28)		moru (A=3,24)	
	f	%	f	%
1	2	8,0	1	3,0
2	5	20,0	10	30,3
3	8	32,0	10	30,3
≥ 4	10	40,0	12	36,4
Yhteensä	25	100	33	100

Piilotermin frekvensseissä ei koehenkilöryhmien välillä ole suuria eroja. Kummankin ryhmän piilotermeistä kaksi kolmasosaa on korkeintaan yksifrekvenssisiä. Kaksi-, kolmi- ja yli nelifrekvenssisiä termejä puuttuu kummankin ryhmän termiehdokaslistoista suunnilleen saman verran. Merenkulun opiskelijoiden termiehdokaslistasta puuttuvat termit esiintyvät tekstissä keskimäärin 2,27 kertaa, kun poimittujen termien keskimääräinen frekvenssi on 4,65. Vähintään kolmen esiintymän termejä puuttuu seitsemän (*бассейновая ЕСКУС, контроль и управление судоходством, морской район А1, морской район А2, район А1, район А2, район А3*). Kieliaineiden opiskelijoiden piilotermit esiintyvät tekstissä keskimäärin vain 1,76 kertaa, mutta piilotermin joukossa on myös kaksi termiä, jotka esiintyvät lähtötekstissä viisi kertaa (*морской район А1, участник транспортного процесса*). Kummallakin koehenkilöryhmällä oli kuitenkin eniten vaikeuksia pienifrekvenssisten termien poiminnassa.

TAULUKKO 23. Venäläisten koehenkilöryhmien piilotermin frekvenssi

Frekvenssi	koru (A=1,76)		moru (A=2,27)	
	f	%	f	%
≤ 1	16	64,0	22	66,7
2	4	16,0	4	12,1
3	1	4,0	2	6,1
≥ 4	4	16,0	5	15,1
Yhteensä	25	100	33	100

Piilotermin pituudessa tai frekvenssissä ei koehenkilöryhmien välillä ollut juuri eroa. Sen sijaan piilotermin termikaavoissa on ryhmien välillä selkeä ero. Venäläisten merenkulun opiskelijoiden termiehdokaslistasta puuttuvien termien yleisimmät termikaavat ovat  $A_xN_x$  (9), jossa x on adjektiivien tai substantiivien lukumäärä ja voi saada arvon

1–3 (*бассейновая информационная система*), ja  $N_x$  (6), jossa  $x$  voi saada arvon 2–4 (*контроль движения судов*). Suluissa on termikaavan mukaisten piilotermin lukumäärä. Lisäksi ehdokkaiden joukosta puuttuu termejä, joissa termin osana on lyhenne (*бассейновая ЕСКУС*), symboli (*морской район А1*), prepositio (*контроль за движением судов*) tai konjunktio (*контроль и управление судоходством*).

Venäläisten kieliaineiden opiskelijoiden termiehdokaslistasta puuttuvien termien yleisimmät termikaavat ovat  $N_x$  (7), jossa  $x$  saa arvon 1–4 (*квалификация экипажа*), prepositiolauseke (*контроль над судоходством*), joka voi alkaa adjektiivilla (7) ja  $A_x$   $N_x$  (4), jossa  $x$  saa arvon 1–3 (*морской флот*). Kieltenopiskelijoiden termiehdokaslistasta puuttuu yksi lyhenne (*GPS*), kaksi termiä, joissa termin osana on symboli (*морской район А1*, *морской район А2*), ja kaksi termiä, joissa termin osana on konjunktio (*контроль и управление судоходством*, *управление и информационное обеспечение судоходства*). Preposition tai konjunktion sisältävät termit ovat pituudeltaan 4–6 sanaa, kun prepositiot ja konjunktiot lasketaan erillisiksi sanoiksi.

Venäläisten kieliaineiden opiskelijoiden piilotermeistä 19 (76 %) on samoja kuin venäläisten merenkulun opiskelijoiden piilotermit (*GPS*, *государство флага*, *контроль за движением судов*, *морской район А1*, *обеспечение безопасного плавания*, *система по надзору за судоходством*, *управление движением судов*). Merenkulun opiskelijoista poiketen kieliaineiden opiskelijat eivät valinneet termejä *бедствие* (häätä, onnettomuus), *квалификация экипажа* (miehistön pätevyys), *контроль над судоходством* (meriliikenteen valvonta), *морской флот* (laivasto), *участник транспортного процесса* (kuljetusketjun lenkki). Listassa ovat kuitenkin termiehdokkaat *экипаж*, *контроль судоходства*, *флот*, *транспортный процесс*. Piilotermit ovat yleiskielen sanoja (*бедствие*, *морской флот*) tai lähikäsitteiden nimityksiä (*квалификация экипажа*, *участник транспортного процесса*). Vain yksi, *контроль над судоходством*, on selkeästi merenkulun termi, jota kielten opiskelijat eivät ole tunnistaneeet. Sen sijaan he ovat kyllä tunnistaneeet variantin *контроль судоходства*.

Merenkulun opiskelijoiden listaan eivät kieliaineiden opiskelijoista poiketen päässeet termit *бассейновая ЕСКУС* (alueellinen merenkulun ohjaus- ja informaatiojärjestelmä), *бассейновая информационная система* (alueellinen informaatiojärjestelmä), *бассейновый информационный центр* (alueellinen informaatiokeskus), *береговая система связи* (rannikkoasemaverkko), *навигационная система* (navigointijärjestelmä), *зона контроля движения судов* (meriliikenteen valvontavyöhyke), *несение вахты* (vahdinpito), *район А1* (merialue A1), *район А2* (merialue A2), *район А3* (merialue A3), *система ОБЧ-радиосвязи* (VHF-radioverkko), *система управления и информационного обеспечения судоходства* (merenkulun ohjaus- ja informaatiojärjestelmä), *спасательно-координационный подцентр* (meripelastuslohkokeskus), *спутниковая навигационная система* (satelliittinavigointijärjestelmä). Termin ensimmäisenä osana olevaa adjektiivimääritettä, kuten esimerkiksi adjektiiveja *бассейновый*, *береговой* tai *навигационный*, ei ole pidetty termiin kuuluvana (ks. viisi ensimmäistä termiä), ei myöskään aluejakoa osoittavia symboleita *А1*, *А2* ja *А3*. Termiksi on valittu *район* (alue) tai *морской район* (merialue). Näiden termien poiminnassa ei merenkulun koulutuksesta ole ollut hyötyä, sillä kaikki luetellut termit ovat merenkulun termejä eivätkä yleiskielen sanoja.

Kummankin venäläisen koehenkilöryhmän termiehdokaslistasta puuttuu eniten vähintään kaksisanaisia pienifrekvenssisiä termejä, jotka noudattavat yleisimpiä termikaavoja. Substantiivien sanaliittoja puuttuu kummankin ryhmän termiehdokaslistasta lähes yhtä monta, mutta adjektiivien ja substantiivien sanaliittojen poiminnassa kieliaineiden opiskelijat onnistuivat paremmin kuin merenkulun opiskelijat.

### 6.5.3 Venäläisten opiskelijaryhmien poimimien valetermien ominaisuuksia

Venäläisten merenkulun opiskelijoiden termiehdokaslistassa on 179 ja kieliaineiden opiskelijoiden listassa 79 sellaista termiehdokasta, joita asiantuntijat eivät valinneet termeiksi. Valetermit ovat keskimääräiseltä pituudeltaan hyvin lähellä koehenkilöiden valitsemien termien pituutta. Venäläisten merenkulun opiskelijoiden valitsevat valetermit ovat keskimäärin 1,76 sanan mittaisia. Puolet näistä termiehdokkaista on yhden sanan ja runsas kolmannes kahden sanan mittaisia. Kolmisanaisia valetermejä on alle 10 prosenttia. Nelisanaisia ja vähintään viisisanaisia valetermejä on kumpiakin neljä. Kieliaineiden opiskelijoiden poimimat valetermit ovat keskimäärin kahden sanan mittaisia. Selkein ero koehenkilöryhmien välillä on yksi- ja kaksisanaisten valetermien poiminnassa. Merenkulun opiskelijat poimivat runsaasti yksisanaisia valetermejä mutta kieliaineiden opiskelijat poimivat eniten kaksisanaisia valetermejä.

TAULUKKO 24. Venäläisten koehenkilöryhmien poimimien valetermien pituus

Sanaluku	koru (A=2,01)		moru (A=1,76)	
	f	%	f	%
1	28	35,5	89	49,7
2	31	39,2	69	38,5
3	15	19,0	13	7,3
≥ 4	5	6,3	8	4,5
Yhteensä	79	100	179	100

Kieliaineiden opiskelijoiden poimimat valetermit esiintyvät lähtötekstissä keskimäärin 2,76 kertaa. Yli puolet valetermeistä on pienifrekvenssisiä ( $f \leq 1$ ) mutta listassa on myös vähintään viisi kertaa tekstissä esiintyviä valetermejä: *ИМО* (5), *Росморфлот* (5), *транспортный процесс* (6), *информационная подсистема* (7), *побережье* (8), *локальная ЕСКУС* (9), *ТОИТЦ* (9), *транспорт* (9), *бассейновый* (10), *залив* (12), *Тихоокеанский бассейн* (12), *ГМССБ* (17). Suluissa on valetermin frekvenssi koetekstissä.

Merenkulun opiskelijat poimivat termiehdokaslistaansa huomattavasti enemmän sekä lyhyitä että pitkiä valetermejä kuin kieliaineiden opiskelijat. 40 prosenttia merenkulun opiskelijoiden valetermeistä esiintyy tekstissä korkeintaan kerran, mutta lähes yhtä moni esiintyy neljästi tai useammin. Tästä syystä valetermien keskimääräinen frekvenssi on lähellä viittä eli kaksinkertainen kieliaineiden opiskelijoiden valetermien frekvenssiin verrattuna. Runsaafrekvenssisten termiehdokkaiden joukossa on lyhenteitä (*ГМССБ*, *ТОИТЦ*), adjektiiveja (*бассейновый* 10, *информационный* 45, *международный* 22, *морской* 27, *региональный* 25) yleiskielen sanoja (*данные* 10, *залив* 12, *информация* 40, *контроль* 22, *обеспечение* 12, *обмен* 15, *основа* 11, *район* 27, *связь* 16, *система* 47, *центр* 14), tietotekniikan alan termejä (*обмен информацией* 10, *информационная система* 12) ja yksi yksilökäsitteen nimi (*Тихоокеанский бассейн* 12). Termiehdokkaan jäljessä oleva numero ilmaisee frekvenssin tekstissä.

TAULUKKO 25. Venäläisten koehenkilöryhmien poimimien valetermien frekvenssi

Frekvenssi	koru (A=2,76)		moru (A=4,91)	
	f	%	f	%
≤ 1	42	53,2	71	39,7
2	12	15,2	26	14,5
3	8	10,1	19	10,6
≥ 4	17	21,5	63	35,2
Yhteensä	79	100	179	100

Venäläisten kieltenopiskelijoiden valitsemien ylimääräisten termiehdokkaiden joukossa on eniten adjektiivin ja substantiivin yhdistelmiä (A N), joita listassa on 16 (20,3 %). Toiseksi suurin ryhmä ovat lyhenteet, joita on 79 valetermin listassa peräti 14. Merenkulun opiskelijoista poiketen kieltenopiskelijat poimivat vain kaksi ylimääräistä adjektiivia (*радиолокационный, бассейновый*). Yleiskielen sanoja on termiehdokaslistassa myös vähemmän (*залив, промысл, транспорт*). Kieltenopiskelijoiden listassa on ylipäätään vähän termiehdokkaita, jotka eivät voisi olla merenkulun käsitteen nimityksiä.

Venäläisten merenkulun opiskelijoiden valetermit noudattavat samoja termikaavoja kuin asiantuntijoiden kanssa samoin poimitut termit. Eniten valetermien listassa on yksisanaisia substantiiveja. Toiseksi eniten on adjektiivin ja substantiivin sanaliittoja, noin neljännes, ja kolmanneksi eniten kahden substantiivin sanaliittoja. Yllättävää on, että valetermien listassa on peräti 10 lyhennettä. Listassa on myös 11 adjektiivia, jotka kyläkin toimivat termin osina, mutta eivät yksinään olleet termejä (*бассейновый, береговой, локальный, международный, морской, региональный, судоходный, транспортный, экономический*), yksilökäsitteiden nimiä (*ИМО, Инмарсат, МКУБ, Тихоокеанский бассейн, Токийский меморандум, Черноморско-Азовское побережье*), yleiskielen sanoja (*администрация, данные, движение, основа, положение, поправка*) ja tietotekniikan termejä (*банк данных, глобальная компьютерная сеть, информационная система*).

Valetermien poiminnan yksimielisyydeksi merenkulun opiskelijoiden ryhmässä saadaan 5,47 (979 jaettuna 179:llä) ja kieltenopiskelijoiden ryhmässä 5,63 (445 jaettuna 79:llä). Keskimäärin viisi tai kuusi opiskelijaa siis valitsi termiehdokkaat, joita asiantuntijat eivät valinneet. Moni merenkulun opiskelija valitsi termiehdokkaat *акватория порта* (10), *континентальный шельф* (10), *Международная ассоциация маячных служб* (10), *техническая характеристика* (10). Suluissa on nämä termiehdokkaat valinneiden opiskelijoiden määrä. Kieliaineiden opiskelijoiden listassa on neljä termiehdokasta, joita jokainen tämän ryhmän koehenkilö piti terminä. Nämä termiehdokkaat ovat *ГМССБ, ОБЧ-радиосвязь, РЛС* ja *ТОИТЦ*. Myös merenkulun opiskelijat pitivät näitä lyhenteitä termeinä, tosin *ОБЧ-радиосвязь* esiintyy merenkulun opiskelijoiden termiehdokaslistassa vain lyhenteenä *ОБЧ*.

## 6.6 Erikoisalatiedon ja terminologian tuntemuksen vaikutus terminpoimintaan

Tutkimukseeni kuuluneessa terminpoimintakokeessa pyysin koehenkilöitä poimimaan tietyn erikoisalan termejä. Kokeessa oli ennakko-oletuksena, että poimijan erikoisalan tai terminologian teorian tuntemus vaikuttaa terminvalintaan. Terminpoiminnan sanotaan perustuvan intuitioon, mutta itse asiassa poimija hyödyntää tehtävän aikana monipuolisesti aivoihin tallennettua maailmantietoa, kielitietoa ja erikoisalatietoa. Tässä jaksossa selostan, miten näistä erilaisista tiedon lajeista juuri erikoisalatieto tai sen puute mahdollisesti vaikutti koehenkilöryhmien terminpoimintaan.

Terminpoimija joutuu poiminnan aikana jokaisen termiehdokkaan kohdalla vastamaan kysymykseen, onko se termi vai yleiskieltä ja kuuluuko se juuri tälle erikoisalalle. Sekä suomen- että venäjänkielisessä tekstissä on sanoja ja sanaliittoja, jotka kuuluvat yleiskieleen, mutta ovat myös merenkulun termejä. Yleiskielen sana tai sanat ovat termejä tai termin osia, jos niillä on erikoisalalla yleiskielisestä merkityksestä poikkeava merkitys joko yksin, sanaliittona tai erikoisalan termiin yhdistyneenä (vrt. Fulford 2001: 261). Esimerkiksi sanat *asema* ja *kulku* tai *бассейн* ja *бедствие* kuuluvat yleiskieleen, mutta ne ovat myös termejä, jotka nimeävät merenkulun käsitteitä. Ahmad ja Rogers (2001: 752) puhuvat kaksitoimisista sanoista (double duty), joilla on yksi tehtävä yleiskielessä ja toinen tehtävä erikoiskielessä (esim. *platform*). Teksteissä on myös sanaliittoja, joiden osat erillisinä ovat yleiskielisiä, mutta yhdessä merenkulun termi. Tällainen sanaliitto on esimerkiksi *vaarallinen aine*, jonka osat erillisinä ovat hyvin tavallisia yleiskielen sanoja, mutta yhdessä ne ovat keskeinen merenkulun turvallisuuden termi. Yleensä kaksiosaisissa termeissä pääsana yksinään kuuluu yleiskieleen, mutta määritteen kanssa siitä tulee erikoisalan termi. Koska monisanaisesta termistä käytetään tekstissä usein lyhyttä muotoa, yleiskielen sana onkin tällöin termin lyhyt muoto (ks. jakso 5.4).

Rajanveto erikoiskielen termien ja yleiskielen sanojen välillä ei ollut helppoa asiantuntijoillekaan. Suomalaisten asiantuntijoiden termilistassa olivat ensimmäisen kierroksen jälkeen esimerkiksi sanat *edusta*, *kulku*, *radio*, *meri*, *merenkulku*, *salmi* ja *tehokkuus*. Näistä termilistaan valittiin lopulta *kulku*, *radio*, *merenkulku* ja *tehokkuus*. Koehenkilöt eivät pitäneet termineinä näitä sanoja, joista vain tehokkuus pääsi opiskelijoiden termiehdokaslistalle. Sitä piti terminä neljä merenkulun opiskelijaa. Suomalaisten koehenkilöiden erikoisalatuntemus tai sen puute näkyy selvimmin termin *vaarallinen aine* poiminnassa. Kääntämisen opiskelijat eivät poimineet tätä sanaliittoa, kun taas kahdeksan merenkulun opiskelijaa poimi. Merenkulun opiskelijat poimivat myös termit *laivan sijainti* ja *sijainti*, joita ei ole kääntämisen opiskelijoiden termiehdokaslistassa. Näiden termien poiminnassa merenkulun opiskelijat eivät kuitenkaan olleet yksimielisiä, sillä termit ovat vain neljän koehenkilön termiehdokaslistassa.

Ennako-oletusten perusteella kääntämisen opiskelijoiden poimintatuloksessa ei olisi pitänyt olla yksilökäsitteitä, koska heidän olisi pitänyt tietää, että termi nimeää erikoisalan yleiskäsitettä. Eräistä merenkulun järjestöistä käytetyt lyhenteet *IALA*, *IMCO* ja *IMO* ovat kuitenkin kummankin suomalaisen koehenkilöryhmän termiehdokaslistassa – kuten myös asiantuntijoiden termilistassa. Termiehdokaslistaan pääsy on hyvin todennäköistä, jos yksilökäsitteen nimitys tai sen osa on lyhenne (*MIF*, *HAZMAT-direktiivi*). Kääntämisen opiskelijat eivät kuitenkaan ole valinneet termeiksi *Englannin kanaalia*, *Elbe-jokea*, *Hampurin satamaa* tai *Malaccan salmea*, jotka 4–7 merenkulun opiskelijaa on valinnut termiehdokaslistaansa. Näiden yksilökäsitteiden nimitysten tunnistamisessa terminologian teorian tuntemuksesta on ollut hyötyä.

Venäläisistä asiantuntijoista jokainen valitsi tavallisimmin yksisanaisia substantiivi-termejä, joilla on myös yleiskielen merkitys, kuten esimerkiksi termeillä *океан* (valtameri), *канал* (kanava), *промысел* (ammatti, elinkeino), mutta valinta ei läheskään aina ollut yksimielinen. Merenkulun erikoiskielen termeiksi kaikki kolme venäläistä asiantuntijaa valitsi yleiskielestä tutut sanat *агент* (agentti, asiamies), *бассейн* (allas, vesialue) ja *бедствие* (häätä, onnettomuus). Myös koehenkilöt valitsivat nämä termit termiehdokaslistoihinsa. Ryhmien välinen ero näkyy termin *бедствие* valinnassa, sillä sitä eivät kieliaineiden opiskelijat pitäneet terminä.

Kysymykseen, onko erikoisalan tuntemuksella vaikutusta terminvalintaan, voidaan etsiä vastausta tarkastelemalla termejä, jotka eri opiskelijaryhmiltä ovat jääneet poimi-



matta. Vertailussa ovat ne termiehdokkaat, jotka valitsi vähintään neljä opiskelijaa kustakin koehenkilöryhmästä.

Suomalaisten kääntämisen opiskelijoiden termiehdokaslistasta puuttuivat muun muassa seuraavat merenkulun opiskelijoiden valitsemat termit (suluissa on niiden merenkulun opiskelijoiden määrä, jotka ovat termin valinneet):

etsintä (5)	pelastus (4)
laivan sijainti (4)	rantavaltio (9)
lastin käsittely (7)	sijainti (4)
näkyvyys (8)	turvallisuus (9)
ohjailla (9)	vaarallinen aine (8)
ohjaus (9)	vesikuljetus (6)
onnettomuus (5)	ympäristöriski (4)

Näiden termien valintaan on merenkulun erikoisalatiedolla ilmeisesti ollut vaikutusta. Mahdollinen selitys on, että osa näistä termeistä on niin yleiskielisiä, että kääntämisen opiskelijat eivät ole pitäneet niitä merenkulun termeinä. Tällaisia ovat esimerkiksi termit *onnettomuus*, *turvallisuus*, *laivan sijainti*, *lastin käsittely* ja *vaarallinen aine*. Niillä on merenkulussa kuitenkin yleiskielen merkityksestä poikkeava käsitesisältö, jolloin ne ovat pettävästä ulkonäöstään huolimatta merenkulun termejä. Voidaan tosin myös väittää, että todellisuudessa termit *onnettomuus* ja *turvallisuus* ovatkin elliptisiä ja täydelliset termimuodot olisivat *merionnettomuus* ja *meriturvallisuus*. Merenkulun opiskelijat ehkä täydentävät elliptiset muodot mielessään, jolloin nämä yleiskieliset sanat nimeävät heidän mielestään koetekstin kontekstissa merenkulun käsitteitä.

Kääntämisen opiskelijoiden termiehdokaslistasta puuttuvien termien joukossa on yksi verbitermi, *ohjailla*. Merenkulun opiskelijat ovat valinneet termin, koska he tietävät, että komentosillalla vahtia ajava perämies nimenomaan ohjailee laivaa. Merenkulun opiskelijoiden termiehdokaslistalla on toinenkin verbitermiehdokas: *purjehtia*. Laivat purjehtivat edelleen, vaikka käyttövoimana tavallisesti on muu kuin tuuli. Tämän termiehdokkaan oli kuitenkin valinnut vain viisi merenkulun ja kaksi kääntämisen opiskelijaa, eivätkä asiantuntijatkaan valinneet sitä termilistaansa.

Merenkulun opiskelijoiden termiehdokaslistasta puuttuivat muun muassa seuraavat kääntämisen opiskelijoiden valitsemat termit (suluissa on niiden kääntämisen opiskelijoiden määrä, jotka ovat termiehdokkaan valinneet):

jakelu (8)	maalle rakennettu tutka-asema (7)
laivakäyttöinen tutka (11)	ohjauspalvelu (5)
laivaliikenteen ohjauspalvelu (9)	varastointi (9)
VTMIS-palvelu (10)	VTMIS toiminta (11)
horisontaalinen integraatio (10)	vertikaalinen integraatio (10)
horisontaalinen tietojenvaihto (9)	vertikaalinen tietojenvaihto (9)

Asiantuntijat ovat pitäneet näitä termiehdokkaita merenkulun termeinä. Näin ollen merenkulun opiskelijoilta on jäänyt poimimatta oman alansa termejä. Yllättävää on, ettei *laivakäyttöinen tutka* ole merenkulkijoiden suosikkitermien joukossa. *Tutkankin* valitsi termiksi vain yhdeksän merenkulun opiskelijaa 11:stä. Ehkä merenkulun opiskelijat ovat tunnistaneet sanaliittoon *laivakäyttöinen tutka* tahattomasti piilotetun ansan, sillä oikea nimitys on *laivatutka*, eikä *laivakäyttöinen tutka* siis tarkkaan ottaen olekaan termi. *VTMIS toiminnan* ja *VTMIS-palvelun* vähäinen suosio merenkulun opiskelijoiden valinnassa johtunee siitä, että termiksi on valittu vain lyhenne *VTMIS* eikä sanojen *toiminta* ja *palvelu* ole katsottu kuuluvan termiin. Asiantuntijat ovat valinneet termilistaan myös nimitykset *horisontaalinen integraatio*, *horisontaalinen tietojenvaihto*, *vertikaalinen integraatio* ja *vertikaalinen tietojenvaihto*, jotka eivät merenkulun opiskelijoiden



mielestä ole merenkulun termejä. He ovat mahdollisesti pitäneet niitä pikemmin lähialan kuin merenkulun termeinä.

Venäläisten kieliaineiden opiskelijoiden termiehdokaslistasta puuttuivat seuraavat venäläisten merenkulun opiskelijoiden valitsemat termit (suluissa on niiden merenkulun opiskelijoiden määrä, jotka ovat termin valinneet):

бедствие (7)

квалификация экипажа (5)

контроль над судоходством (6)

морской флот (4)

участник транспортного процесса (8)

Termin *бедствие* (häätä) valintaan on merenkulun erikoisalatiedolla ilmeisesti ollut vaikutusta, sillä noin kaksi kolmasosaa merenkulun opiskelijoista on pitänyt sanaa terminä. Muiden termien valintaan ovat vaikuttaneet muut syyt kuin merenkulun erikoisalatieto, sillä kieliaineiden opiskelijoiden listassa ovat yläkäsitteitä nimeävät termit *флот* ja *экипаж*, termiehdokas *транспортный процесс* ja termivariantti *контроль судоходства*.

Venäläisten merenkulun opiskelijoiden termiehdokaslistasta puuttuivat seuraavat kieliaineiden opiskelijoiden valitsemat termit (suluissa on niiden kieliaineiden opiskelijoiden määrä, jotka ovat termiehdokkaan valinneet):

бассейновая ЕСКУС (9)

район А1 (6)

бассейновая информационная система (7)

район А1 (6)

береговая система связи (8)

район А3 (6)

зона контроля движения судов (4)

система ОБЧ-радиосвязи (4)

навигационная система (6)

несение вахты (5)

спутниковая навигационная система (4)

спасательно-координационный подцентр (5)

система управления и информационного обеспечения судоходства (6)

Suurin osa näistä termeistä liittyy kansipäällystön jokapäiväisiin työtehtäviin, joten niiden valitsematta jättämistä on vaikea selittää. Ellei syy olekin juuri se, että ne ovat liian arkisia. Toisaalta merenkulun opiskelijoilta poimimatta jääneiden mutta kieliaineiden opiskelijoiden poimimien ja toisaalta kieliaineiden opiskelijoilta poimimatta jääneiden mutta merenkulun opiskelijoiden poimimien termien vertailun perusteella terminologian teorian tuntemuksesta on terminpoiminnassa enemmän hyötyä kuin erikoisalan tuntemuksesta.

Suomalaisten ja venäläisten koehenkilöryhmien termiehdokaslistojen vertailu osoittaa, että koehenkilöt poimivat termejä varsin samalla tavalla käsitelähtöisesti koulutuksesta riippumatta. Tulos tukee Kobrinin (1989: 23, 27) väitettä, jonka mukaan kielenpuhujat tunnistavat termit helposti tekstistä eikä tulos riipu erilaisista koulutustaustoista, sillä termit tunnistetaan ja valitaan ilmausten kieliopillisten ja semanttisten ominaisuuksien perusteella. Koulutuksen sijasta terminpoimintaan vaikuttavat pikemminkin yksilölliset käsitykset siitä, mikä on termi.

## 6.7 Puoliautomaattinen terminpoiminta

Edellä olen selostanut koetta, jossa termit poimittiin käsin samaa aihetta käsittelevästä suomenkielisestä ja venäjänkielisestä noin 2 500 sanan tekstistä. Manuaalinen terminpoimintamenetelmä on edelleen toimiva ja paljon käytetty. Tietotekniikka antaa kuitenkin mahdollisuuden automatisoida tätä aikaa vievää sanastotyön vaihetta. Termintunnistushjelmia onkin kehitetty 1980-luvulta lähtien. Koska tässä tutkimuksessa yhtenä tavoitteena oli selvittää puoliautomaattisen terminpoiminnan käyttökelpoisuutta, termit

tunnistettiin samoista koeteksteistä manuaalisen poiminnan lisäksi kahdella kaupallisella termintunnistusohjelmalla. Termiehdokkaat poimittiin sekä ohjelmien perusasetuksilla että asetuksia muuttamalla. Tavoitteena oli löytää ohjelman asetuksista sellainen yhdistelmä, jolla puoliautomaattisen poiminnan tulos olisi mahdollisimman lähellä asiantuntijoiden tekemän manuaalisen poiminnan tulosta. Vertasin eri asetuksilla tuotettuja termiehdokaslistoja asiantuntijoiden termilistoihin, ja valitsin varsinaiseen vertailuun ne termiehdokaslistat, joissa saanti oli korkein, eli termiehdokaslistat, joissa oli eniten termejä.

### **6.7.1 Testattavien terminpoimintaohjelmien esittely**

Suomenkielisestä tekstistä termit tunnistettiin Conexor Oy:n NaviTerm 2.0 ja Trados GmbH:n MultiTerm Extract -termintunnistusohjelmalla. Venäjänkielisestä tekstistä termit tunnistettiin vain MultiTerm Extract -ohjelmalla, koska NaviTerm 2.0 -ohjelma ei toistaiseksi pysty tunnistamaan venäjänkielisiä termejä. Lisäksi suomenkieliset termit tunnistettiin Master's Innovations Oy:n ohjelmalla, josta ei tutkimukseni aikaan vielä ollut kaupallista versiota. Tästä syystä en esittelen ohjelmaa tarkemmin. Seuraavaksi selostan lyhyesti tässä tutkimuksessa käytettyjen kaupallisten termintunnistusohjelmien ominaisuuksia.

#### **6.7.1.1 Conexor NaviTerm 2.0**

NaviTerm 2.0 -ohjelmaa markkinoivan Conexor Oy:n yleisesitteen mukaan ohjelma on tarkoitettu automaattiseen indeksointiin, asiasanojen ja erisnimien tunnistukseen sekä teknisten termien tunnistukseen. NaviTerm 2.0 -ohjelma on eräänlainen automaattinen indeksioija, joka uuttaa asiakirjan sisältöä parhaiten kuvaavat yksisanaiset ja monisanaiset indeksitermit (Lahtinen 2000: 14). Ohjelmaa ei siis ole kehitetty erityisesti teknisten termien eli terminologian teorian mukaisten termien tunnistamiseen. Koska indeksitermejä on teksteissä yleensä enemmän kuin teknisiä termejä (Lahtinen 2000: 34), NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa on jo tällä perusteella runsaasti roskaa, jos tavoite on tuottaa vain teknisiä termejä sisältävä lista (ks. jakso 2.3.1).

NaviTerm 2.0 -ohjelma tekee tekstille sisällönanalyysin, johon kuuluvat sanaluokantunnistus, leksikaalinen ja syntaktinen analyysi sekä sanojen sijainnin ja jakauman analyysi. Yrityksen esitteen mukaan sisällönanalyysi ulottuu koko tekstiin eikä vain lause-tasolle. Lisäksi ohjelma käyttää termiehdokkaiden valinnassa tilastollista valintaperustetta. Termintunnistus perustuu ajatukseen, että termeillä on samanlaisia morfologisia, syntaktisia ja leksikaalisia ominaisuuksia ja että jotkin kielelliset elementit tai sanat ovat hyviä ja jotkin huonoja termin osina, esimerkiksi pääsanat attribuutteina. NaviTerm 2.0 on opetettu tunnistamaan termit tekstistä termikaavojen avulla. Ohjelma hakeekin tekstistä substantiivilausekkeita, joiden ihannepituus on 1–3 sanaa. Sulkulistaa ei ohjelmassa ole valmiina, mutta halutessaan käyttäjä voi sen itse lisätä (Timo Lahtinen sähköpostitse 1.3.2004). Ohjelman käyttämä termintunnistusalgoritmi on raportoitu Timo Lahtisen (2000) väitöskirjassa.

Lingvistisen analyysin vuoksi NaviTerm 2.0 -ohjelma on kielestä riippuva. Kielivalikoimassa oli terminpoimintakokeen tekohetkellä suomi ja englanti. Terminhaun lähteeksi voidaan valita monta asiakirjaa, joiden tallennusmuoto voi olla esimerkiksi HTML tai teksti. Käyttäjä voi muokata (customize) ohjelman tukemaan myös muita tiedostomuotoja. NaviTerm 2.0 -ohjelman kehittäjien mukaan tekstin vähimmäispituus on noin 10 sivua (Timo Lahtinen suullisesti 19.9.2003).

Sisällönanalyysin perusteella termiehdokkaat pisteytetään termimäisyyden mukaan lähes portaattomasti. Paras termiehdokas saa pistemäärän 1.0 ja muut ehdokkaat 0.xxxx. Moni termi voi saada saman pistemäärän. Pisteytyksen periaatteet on esitelty Lahtisen (2000) väitöskirjatutkimuksessa. Pisteytyksessä käytetään muun muassa ennalta suures- ta korpuksesta laskettua painokerrointa (Timo Lahtinen sähköpostitse 1.3.2004) ja asia- kirjakohtaista purskeisuutta.

Indeksitermien tunnistuksessa termiehdokkaiden pisteytys on hyvinkin perusteltua, koska tavoitteena on löytää asiakirjan sisältöä parhaiten kuvaavat ilmaisut. Terminolo- giassa termit ovat periaatteessa samanarvoisia ja teksteistä pyritään tunnistamaan termit ja vain termit mahdollisimman kattavasti. Voidaan kuitenkin olettaa, että korkeita pis- temääriä saaneet termit nimeävät todennäköisesti keskeisiä käsitteitä, jotka sijoittuvat käsitejärjestelmässä lähelle huippua.

NaviTerm 2.0 -ohjelma tuottaa listan perusmuotoon palautetuista termiehdokkaista noin kolmessa minuutissa, jos tekstin pituus on korkeintaan noin 11 000 sanaa. Listassa on eroteltu tekniset termit, indeksitermit sekä järjestöjen, paikkojen ja henkilöiden ni- met. Listan voi järjestää kuudella eri tavalla. Mahdollisia tapoja ovat aakkosten mukaan järjestäminen joko nousevasti tai laskevasti, termiyspisteiden mukaan nousevasti tai laskevasti tai esiintymien lukumäärän mukaan nousevasti tai laskevasti. Termiehdokkai- ta voi poistaa ja lisätä ja niiden kirjoitusasua voi korjata. Termiehdokkaiden määrää voi vähentää liukukytkimellä, jonka avulla valitaan näkymään tietty prosenttiosuus ter- miehdokkaista. Näytöllä on myös valitun lähdeasiakirjan minikuva, jossa valitun termin esiintymät on osoitettu punaisilla neliöillä. Valitsemalla jokin neliöistä saadaan näky- viin termin tekstikonteksti.

Termiehdokaslista on mahdollista tulostaa vain aakkosjärjestyksessä, eivätkä ter- miyspisteet tai esiintymien lukumäärät siirry tulosteeseen automaattisesti. Ohjelma ei ole erehtymätön termiehdokkaiden luokittelussa teknisiin termeihin, indeksitermeihin sekä järjestöjen, paikkojen ja henkilöiden nimiin. Jos käyttäjä haluaa tarkastella esimer- kiksi vain teknisiä termejä, luokittelu on syytä tarkistaa manuaalisesti.

#### 6.7.1.2 Trados MultiTerm Extract

MultiTerm Extract -termintunnistusohjelmaa markkinoiva yritys pitää ohjelman termin- tunnistusperiaatteet liikesalaisuutena, mutta termiehdokaslistan perusteella MultiTerm Extract -ohjelma käyttää termintunnistuksessa tilastollisen ja jonkinlaisen lingvistisen menetelmän yhdistelmää. Hallitseva tunnistusmenetelmä on kuitenkin sanojen ja sa- nayhdistelmien frekvensseihin perustuva tilastollinen menetelmä. Ohjelma tuottaa lis- tan, jossa termiehdokkaat on järjestetty termiyspisteiden mukaan. Parhaat termiehdok- kaat saavat pistemäärän 99. Pisteet ovat aina kokonaislukuja, joten samalla pistemääräl- lä on varsinkin listan loppupäässä jopa satoja termiehdokkaita.

Periaatteessa ohjelmassa on toiminto, jolla termiehdokkaat palautetaan perusmuo- toon, mutta ainakaan suomen ja venäjän kielen osalta tämä toiminto ei terminpoiminnan tekohetkellä ollut vielä täysin kehittynyt. Ohjelma tunnistaa eri taivutusmuodoissa ole- via substantiiveja samaksi termiksi, mutta termiehdokaslistaan saattaa silti tulostua usei- ta saman kantasanen taivutusmuotoja. Esimerkiksi *alueellisen* ja *alueellisia* ovat ter- miehdokaslistassa erikseen, samoin *kansallista* ja *kansallisten* sekä *käyttö*, *käyttöön*, *käytöstä* jne.

Termiehdokaslistaa voi muokata, ja termiehdokkaita voi poistaa, lisätä tai muuttaa. Ohjelma myös näyttää termiehdokkaan kontekstin. Perusoletuksena on kielikohtaisen sulkulistan käyttö, ja ohjelmaan on tallennettu muun muassa suomen ja venäjän sulkul- listat. Suomen sulkulistassa on 398 sanaa, jotka ovat sanaluokaltaan tavallisesti pronon-

mineja, adverbbeja tai verbejä. Lista on aivan liian suppea, jotta se toimisi tehokkaasti, koska jo yhden verbin koko paradigma käsittää kymmeniä sanoja. Venäjän sulkulistassa on 587 sanaa (pronomineja, prepositioita, konjunktioita ja adjektiiveja). Koska myös venäjän sanat taipuvat, ei venäjän sulkulista ole tehokkaampi kuin suomenkaan sulkulista. Venäjän sulkulistassa ovat muun muassa adjektiivit *большой* (suuri), *хороший* (hyvä) ja *плохой* (huono). Näin lista saattaa jopa estää oikeiden termien valinnan, sillä esimerkiksi termissä *хорошая морская практика* (hyvä merimiestapa) adjektiivi *хороший* on oleellinen osa termiä.

MultiTerm Extract -ohjelma tunnistaa kaikki tavallisimmat tiedostomuodot ja myös muut kuin latinalaiset kirjoitusmerkit. Termintunnistus noin 2 500 sanan suomen- tai venäjänkielisestä tekstistä tapahtuu muutamassa sekunnissa. Tulostus on helppoa tekstinkäsittelyohjelman avulla. Termiehdokkaiden saamat pisteet eivät tässäkään ohjelmassa tulostu. Perusasetuksissa termiehdokkaan pituus on 1–10 sanaa, mutta pituutta voi säätää. Valittavien termiehdokkaiden lukumäärälle voi asettaa ylärajan. Myös roska määrää voi säätää liukupainikkeella, mikä sinänsä on hupaisaa, sillä tavoitteenahan on tietysti lista, jossa ei ole roskaa.

TAULUKKO 26. Yhteenveto vertailussa käytettyjen termintunnistusohjelmien ominaisuuksista

	NaviTerm 2.0	MultiTerm Extract
Termintunnistusmenetelmä	Yhdistelmämenetelmä	Tilastollinen menetelmä pääasiallisesti
Lemmatisointi	on	osittain
Termiehdokkaiden pisteytys	0.xxx–1.0	0–99
Termiehdokkaiden pituus	säädettävissä	säädettävissä
Termiehdokaslistan muokausmahdollisuus	on	on
Kontekstin näyttö	on	on
Sulkulista	ei	on
Termiehdokkaiden lukumäärä	säädettävissä	säädettävissä
Tulostus	aakkosjärjestys	aakkosjärjestys
Kielet	englanti, suomi	37 kieltä, mm. englanti, suomi ja venäjä
Lähtötekstin tiedostomuoto	tavallisimmat tiedostomuodot	tavallisimmat tiedostomuodot
Lähtötekstin pituus	ei rajoitusta	minimi noin 10 sivua
Termintunnistukseen kuluva aika	pari minuuttia	muutama sekunti

### 6.7.2 Terminpoimintaohjelmien asetukset terminpoimintakokeessa

Tutkimukseni terminpoimintakokeessa vertasin puoliautomaattisen termintunnistuksen tuloksia käsinpoiminnan tuloksiin. Tavoitteena oli vastata kysymykseen, miten hyvin kaupalliset termintunnistusohjelmat tunnistavat termit. Arviointiperusteena oli ohjelmien tuottamien termiehdokaslistojen saanti ja tarkkuus verrattuina terminologian ja me-

renkulun asiantuntijoiden käsin poimimien termien listaan. Tein koepoimintoja sekä pitkillä että lyhyillä teksteillä ja säädin ohjelmien perusasetuksia niin, että saanti oli mahdollisimman hyvä.

#### 6.7.2.1 Lähdetekstin pituuden vaikutus termiehdokkaiden määrään puoliautomaattisessa terminpoiminnassa

Tutkimukseni terminpoimintakokeessa käytettyjen termintunnistusohjelmien markkinoijat eivät anna suosituksia käsiteltävän tekstin pituudesta. Koska puoliautomaattiset termintunnistusohjelmat käyttävät yhtenä termiuden kriteerinä esiintymien määrää lähtötekstissä, hyvin lyhyet (sivun tai kahden sivun) lähtötekstit eivät tuota kelvollista termiehdokaslistaa. Tosin tällaisessa tilanteessa ei kannata antaa termintunnistusta epäluotettavan ohjelman tehtäväksi, sillä manuaalinen terminpoiminta parin sivun tekstistä käy varsin nopeasti. Toisaalta ohjelmien suorituskyky näyttää loppuvan kesken, jos lähtötekstin pituus on kymmeniä tuhansia sanoja. Ennen terminpoimintaa halusinkin selvittää lähtötekstin pituuden ja NaviTerm 2.0 -ohjelman tuottamien termiehdokkaiden määrän välisen yhteyden. Valitsin yhdestä merenkulun turvallisuustekstistä (kiv97) tarkasteluun erimittaisia otteita. Seuraavassa taulukossa on esitetty tekstiotteen pituus sanoina ja vastaava termiehdokkaiden määrä.

TAULUKKO 27. NaviTerm 2.0 -ohjelman poimimien termiehdokkaiden määrän riippuvuus lähtötekstin pituudesta

Tekstiotteen pituus	Termiehdokkaiden lukumäärä	Termiehdokkaat/sivu
1 200 sanaa (5 sivua)	479	96
2 941 sanaa (10 sivua)	972	97
6 743 sanaa (21 sivua)	1 859	88-89
11 000 sanaa (33 sivua)	2 894	88
14 743 sanaa (45 sivua)	2 967	66
18 500 sanaa (50 sivua)	2 964	60

Edellä olevasta taulukosta ilmenee, että termiehdokkaiden lukumäärän kasvu hidastuu huomattavasti, kun tekstin pituus lähenee 10 000 sanaa, ja loppuu kokonaan, kun tekstiotteen pituus on noin 15 000 sanaa. NaviTerm 2.0 -ohjelma tuottaakin eniten termiehdokkaita sivua kohti, kun lähtötekstin pituus on noin 3 000 sanaa. Taulukon perusteella 2 000–3 000 sanan mittainen teksti sopii hyvin koetekstiksi, varsinkin jos tavoitteena on hakea tekstistä mahdollisimman paljon termiehdokkaita suhteessa tekstin pituuteen. Toisaalta taulukosta voi päätellä, että jo noin 1 000 sanan mittainen teksti riittää hyvin tuottamaan runsaasti termiehdokkaita. Itse asiassa noin 1 000 sanan teksti olisi todennäköisesti riittänyt myös käsinpoiminnan koetekstiksi, jolloin koehenkilöiden tehtävä olisi ollut helpompi. Lisäksi poimintatulosten kokoaminen, käsittely ja vertailu olisi helpotunut. Esimerkiksi Fulford (2001: 271) käytti käsinpoimintakokeissa kolmen erikoisanan tekstejä, joiden sanaluvut olivat 351, 476 ja 722.

### 6.7.2.2 Terminpoimintaohjelmien asetusten optimointi

NaviTerm 2.0 -termintunnistusohjelmalla termiehdokaslistan voi järjestää aakkosten, frekvenssien tai termiyspisteiden mukaiseen järjestykseen. Ohjelma myös luokittelee termiehdokkaat teknisiin termeihin, indeksitermeihin sekä organisaatioiden, paikkojen ja henkilöiden nimiin. Tämän tutkimuksen vertailuun olisi siis voitu valita listasta vain tekniset termit. Olen kuitenkin käyttänyt vertailussa täydellistä termiehdokaslistaa, koska myös asiantuntijat ovat valinneet erisnimiä eli yksilökäsitteiden nimityksiä, vaikka ne eivät termin määritelmän mukaan olekaan termejä (ISO 1087-1:2000(E/F): 6). Toisen syy täydellisen listan käyttöön on, että ohjelman tekemä luokittelu osoittautui epäluotettavaksi.

Tuotin MultiTerm Extract -ohjelmalla tekstistä termiehdokaslistan sekä ohjelman perusasetuksilla että säätämällä joko roskan määrää tai termin enimmäispituutta sanojen lukumäärällä mitattuna. En rajoittanut termiehdokkaiden lukumäärää tekstin lyhyden vuoksi ja myös siksi, että tässä työssä pyrin tuottamaan erilaisilla terminpoimintamenetelmillä mahdollisimman kattavan termiehdokaslistan. Tällöin vain ohjelman parhaiksi arvioimat termit eivät riitä. Termien enimmäispituuden vaikutusta termiehdokaslistaan tarkastelin oletusarvolla 10 sanaa, yleensä terminologiassa ylärajana pidetyllä arvolla neljä sanaa ja enimmäispituudella kuusi sanaa, joka näyttäisi tutkimukseni käsinpoiminnan tulosten perusteella olevan suomenkielisten termien enimmäispituus (ks. jakso 6.3.1).

MultiTerm Extract -terminpoimintaohjelmassa roskan määrää säättävä liukupainike on perusasetuksissa kohdassa 50 prosenttia. Liukupainikkeen eri asennoilla saanti oli korkein, kun liukupainike oli kohdassa 75 prosenttia roskan puolella. Mitä enemmän liukupainiketta siirretään roskan puolelle, sitä pitempiä termiehdokkaita ohjelma valitsee. Seuraavassa taulukossa on vertailun vuoksi termintunnistuksen tuloksia ohjelman eri säädöillä. Termiehdokkaan vähimmäispituus on yksi sana kaikissa yhdistelmissä. Ohjelman oman sulkulistan käyttö on perusoletuksena.

TAULUKKO 28. MultiTerm Extract -ohjelman säätöjen vaikutus termintunnistukseen lähtötekstillä pw02 (2 366 sanaa)

Termiehdokkaan enimmäispituus	Roskan määrä %	Termiehdokkaiden lukumäärä	Huomautuksia
10	25	20	Termiehdokkaiden lukumäärä on liian alhainen.
10	50	134	Termiehdokkaiden lukumäärä on liian alhainen.
4	50	135	Termiehdokkaiden lukumäärä on liian alhainen.
10	75	306	Termiehdokaslistat ovat täysin identtiset enimmäispituudella 6 tai 10 sanaa.
6	75	306	
10	80	540	Parhaimmiksi arvioidut termiehdokkaat ovat 10 sanan tekstiotteita.
10	100	540	Lähes 50 % termiehdokkaista on 10 sanan tekstiotteita.
6	100	789	Yli puolet termiehdokkaista on 6 sanan tekstiotteita, noin 10 % on yhden tai kahden sanan termiehdokkaita.



Kokeessa termiehdokaslistat tuotettiin MultiTerm Extract -ohjelmalla roskan määrää säättävän liukukytkimen arvoilla 25, 50, 75, 80 ja 100 prosenttia. Kokeen tulosten perusteella voidaan todeta, että MultiTerm Extract -ohjelmassa roskan määrää säätelevää liukukytöntä ei kannata asettaa alle oletusarvon, koska oletusarvoa pienemmillä liukukytkimien asennoilla tunnistettujen termiehdokkaiden lukumäärä jää kovin alhaiseksi. Koska näissäkin listoissa osa termiehdokkaista on valetermiä, oikeiden termien määrä jää erittäin pieneksi.

Roskan määrän rajoittaminen 75 prosenttiin tuotti termiehdokaslistan, jossa on 306 termiehdokasta, kun taas liukukytkimien asettaminen 50 prosenttiin tuotti vain 134 termiehdokasta. Termin enimmäispituuden rajoittamisella ei näillä valetermien määrällä näytä olevan ollenkaan vaikutusta. Kun liukukytöntä säädettiin niin, että valetermien määrä nousi ensin 80 prosenttiin ja sitten 100 prosenttiin, termiehdokaslistat muuttuivat täysin, sillä ohjelma tunnisti parhaiksi termeiksi termille asetetun enimmäispituuden mittaisia tekstiotteita. Tällaisten termiehdokaslistojen manuaalinen käsittely vaatisi niin paljon aikaa, ettei niiden tuottaminen puoliautomaattisesti ole tarkoituksenmukaista.

Koe osoitti, että käyttökelpoisia MultiTerm Extract -ohjelman termiehdokaslistoista ovat perusasetuksilla tuotettu lista ja lista, jossa on muutettu ainoastaan valetermien määrää. Asiantuntijoiden tekemän manuaalisen poiminnan tulokseen verrattuna saanti on korkein, kun liukukytin on kohdassa 75 prosenttia. Valetermien määrän kasvattaminen tätä suuremmaksi ei ole tarkoituksenmukaista, sillä vaikka suuremmalla arvolla saadaan enemmän termiehdokkaita, listassa ei välttämättä ole enempää termejä kuin perusasetuksilla tuotetussa listassa. Lähtötekstistä *pw02* ohjelma tuotti näillä asetuksilla 306 termiehdokkaan listan, jossa sanoja ei ollut manuaalisesti palautettu oikeisiin perusmuotoihin. Kun vertailuun otettiin sama kantasana vain kerran riippumatta siitä, kuinka monessa eri taivutusmuodossa sana esiintyi, termiehdokaslistaan jäi 255 termiehdokasta.

### **6.7.3 NaviTerm 2.0 ja Masterin-ohjelmien tuottaman termiehdokaslistan ja asiantuntijalistan vertailu**

NaviTerm 2.0 -terminpoimintaohjelma tuotti suomenkielisestä tekstistä *pw02* termiehdokaslistan, jossa on 694 ehdokasta. Listassa on 129 sellaista termiehdokasta, jotka myös vertailulistan koostaneet terminologian ja merenkulun asiantuntijat olivat valinneet. Listasta puuttuu 59 asiantuntijoiden valitsemaa termiä. Kun tunnistettujen termien määrä 129 jaetaan asiantuntijoiden valitsemien termien määrällä 189, saadaan ohjelman tuottaman termiehdokaslistan saanniksi 68,3 prosenttia.

NaviTerm-ohjelman lisäksi termit poimittiin suomenkielisestä lähtötekstissä Masterin-terminpoimintaohjelmalla. NaviTerm-ohjelman tavoin tämäkin ohjelma soveltaa lingvististä terminpoimintamenetelmää. Ohjelmien termintunnistusalgoritmit eroavat vain vähän, sillä Masterin-ohjelman tuottamassa termiehdokaslistassa on 690 termiehdokasta, joista termejä on 116. Näistä 99 myös NaviTerm-ohjelma tunnisti termeiksi. Masterin-ohjelman saanti jää hieman alhaisemmaksi, 61,4 prosenttiin.

Jos terminpoiminnassa pyritään mahdollisimman hyvään saantiin eli pyritään poimimaan niin suuri osa lähtötekstin termeistä kuin mahdollista, poimintatuloksen tarkkuus kärsii. Poimintatuloksen alhainen tarkkuus merkitsee valetermien määrän kasvua. Tässä kokeessa NaviTerm 2.0 -ohjelmalla saavutettiin liki 70 prosentin saanti todellakin tarkkuuden kustannuksella, sillä tarkkuus jää 18,6 prosenttiin (oikeat termit/kaikki termiehdokkaat). 694 termiehdokkaan listassa on 565 sellaista termiehdokasta, joita asian-

tuntijat eivät olleet valinneet. Masterin-ohjelman termiehdokaslistassa on valetermejä suunnilleen sama määrä eli 574, jolloin ohjelman tarkkuus on 16,8 prosenttia.

Seuraavaksi vertaan NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien tuottamia termiehdokaslistoja suomalaisten asiantuntijoiden termilistaan. Ensin pyrin vastaamaan kysymykseen, millaisia ovat termit, jotka NaviTerm 2.0 tai Masterin-ohjelma on tunnistanut. Tämän jälkeen tarkastelen termejä, joita ohjelmat eivät tunnistaneet. Lopuksi selvitan, millaisia ovat terminpoimintaohjelmien valitsevat termiehdokkaat, jotka eivät asiantuntijoiden mielestä ole termejä. Vertaan kaikissa tarkasteluissa asiantuntijalistan ja terminpoimintaohjelmien tuottamien termiehdokaslistojen termien ja termiehdokkaiden sanalukua, frekvenssiä ja termikaavoja. Koska tässä tutkimuksessa käytetyt terminpoimintaohjelmat pisteyttävät termiehdokkaat termiiden mukaan, tarkastelen myös termiehdokkaiden saamia termiypisteitä.

#### 6.7.3.1 NaviTerm 2.0 ja Masterin-ohjelmien poimimien termien ominaisuuksia

NaviTerm 2.0 -terminpoimintaohjelman suomenkielisestä artikkelista *pw02* tuottamassa termiehdokaslistassa on 129 sellaista termiä, jotka asiantuntijat valitsivat termilistaansa. Näiden termien keskimääräinen sanaluku on 1,29. Se on hieman alhaisempi kuin asiantuntijoiden valitsemien kaikkien termien sanaluku 1,54, joten terminpoimintaohjelman tunnistamien termien joukossa on suhteellisesti enemmän lyhyitä kuin pitkiä termejä. Yksisanaisten termiehdokkaiden osuus ohjelman tunnistamista termeistä on 76,0 prosenttia, kun asiantuntijoiden listassa yksisanaista termejä on vain 61,4 prosenttia. Ohjelma tunnistikin hyvin yksisanaiset termit, sillä asiantuntijoiden valitsemista 116 yksisanaista termistä NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa on 98 (84,5 %). Monisanaisten termien tunnistus ei onnistunut yhtä hyvin, sillä 54 kaksisanaista termistä ohjelma tunnisti vain puolet, 12 kolmisanaista termistä kolmasosan ja seitsemästä yli nelisanaista termistä ohjelma tunnisti yhden englanninkielisen (*Regional Traffic Information Service*). Heikompi tulos monisanaisten termien tunnistuksessa ei kuitenkaan heikennä saantia merkittävästi, sillä monisanaisten osuus termeistä on vain noin kolmannes (38,6 %).

Masterin-ohjelman tunnistamien termien pituus on lähempänä asiantuntijoiden poimimien termien pituutta, sillä ohjelman poimimista termeistä yksisanaista on 59,5 prosenttia. Ohjelma tunnisti melko tasaisesti eripituisia termejä, sillä asiantuntijoiden valitsemista yksisanaista termeistä ohjelma tunnisti 59,5 prosenttia, kaksisanaista 68,5 prosenttia ja vähintään kolmisanaista noin puolet.

TAULUKKO 29. NaviTerm 2.0 ja Masterin -terminpoimintaohjelmien poimimien termien pituus

Sanaluku	NaviTerm 2.0 (A=1,29)		Masterin (A=1,55)	
	f	%	f	%
1	98	76,0	69	59,5
2	26	20,1	37	31,9
3	4	3,1	7	6,0
≥ 4	1	0,8	3	2,6
Yhteensä	129	100	116	100

NaviTerm 2.0 -ohjelman ja asiantuntijoiden samoin valitsemien termien keskimääräinen frekvenssi on 2,98 eli jonkin verran korkeampi kuin asiantuntijoiden valitsemien termi-

en frekvenssi 2,34. Mitä suurempi frekvenssi termillä on sitä suurempi mahdollisuus sillä on päästä NaviTerm 2.0 -ohjelman termiehdokaslistaan (vrt. Pasanen 2005b). Kriittinen raja näyttää olevan kaksi esiintymää lähtötekstissä. Kuten voi olettaa, ohjelman tuottamasta termiehdokaslistasta puuttuvat kaikki ne asiantuntijoiden valitsevat 15 termiä, joilla on lähtötekstissä nolلافrekvenssi. Vain kerran lähtötekstissä esiintyvien termien tunnistamisesta ohjelma ei suoriudu aivan yhtä hyvin kuin yksisanaisten termien tunnistamisesta, sillä asiantuntijoiden valitsemista 109:sta yhden esiintymän termistä ohjelma tunnisti 73 eli 67 prosenttia. Tunnistettujen termien osuus kasvaa tasaisesti frekvenssin kasvaessa. Kaksi kertaa lähtötekstissä esiintyvistä termeistä NaviTerm 2.0 -ohjelma tunnisti 80 prosenttia ja vähintään kolme kertaa esiintyvistä kaikki kahta termiä lukuun ottamatta. Termit *VTS komitea* ja *VTS toiminta* saattoivat jäädä tunnistamatta siksi, että ne ovat tekstissä sanaliittoina ilman yhdysmerkkiä.

Masterin-ohjelman tunnistamien termien keskimääräinen frekvenssi on 3,15 eli hie-  
man korkeampi kuin NaviTerm-ohjelman tunnistamien termien keskimääräinen frek-  
venssi 2,98. Ohjelman heikko kohta näyttäisi olevan pienifrekvenssisten termien tunnis-  
tus, sillä ohjelma tunnisti korkeintaan kerran lähtötekstissä esiintyvistä termeistä 62 eli  
11 vähemmän kuin NaviTerm 2.0 -ohjelma. Vähintään kaksi kertaa esiintyvien termien  
tunnistuksessa ei ohjelmien välillä ole juuri eroa, kun verrataan tunnistettujen termien  
frekvenssejä.

TAULUKKO 30. NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien poimimien ter-  
mien frekvenssi

Frekvenssi	NaviTerm 2.0 (A=2,98)		Masterin (A=3,15)	
	f	%	f	%
≤ 1	73	56,6	62	53,4
2	28	21,7	27	23,3
3	5	3,9	6	5,2
≥ 4	23	17,8	21	18,1
Yhteensä	129	100	116	100

NaviTerm 2.0 -ohjelman tunnistamista termeistä suurin osa on yhdyssubstantiiveja ( $N_c$ ) samoin kuin asiantuntijoidenkin poimimista termeistä. Ohjelma tunnisti yksisanaista substantiivitermeistä 83 prosenttia. Näiden termien korkea saanti oli odotettavissa, kos-  
ka ohjelma tunnistaa hyvin yksisanaista termejä, kuten totesin edellä. Kahden substan-  
tiivin sanaliittotermeistä ohjelma tunnisti 60,1 prosenttia. Adjektiivin ja substantiivin  
sanaliittojen tunnistamisessa saanti on korkeampi, 68,8 prosenttia. Vaikeuksia ohjelma-  
lla oli englanninkielisen lyhenteen ja suomenkielisen substantiivin muodostamien termi-  
en tunnistamisessa, sillä tämän termikaavan mukaisista 24 termistä ohjelma tunnisti  
vain 10. Syy saattaa olla joidenkin termien kirjoitusasu, sillä kirjoittaja on käyttänyt  
yhdysmerkkiä vain osassa tämän ryhmän termeistä. Englanninkielisistä lyhennetermeis-  
tä ohjelma tunnisti kaikki kymmenen ja englanninkielisistä sanaliittotermeistä joka toi-  
sen. (Vrt. Pasanen 2005b taulukko 5:2.)

Masterin ja NaviTerm 2.0 -ohjelmien poimimien termien sanaluvuissa tai frekvens-  
seissä ei ole selkeää eroa, mutta termikaavoissa eroa on. Suurin osa termeistä, jotka Na-  
viTerm-ohjelma tunnisti hyvin mutta Masterin-ohjelma huonosti, on yhdyssubstantiive-  
ja. Näistä Masterin-ohjelma tunnisti vain 57,6 prosenttia. Toisaalta Masterin-ohjelma  
tunnisti kaikki englanninkieliset termit ja yhdeksän niistä 13:sta lyhenteen ja substantii-  
vin muodostamasta termistä, joissa ei ole käytetty yhdysmerkkiä.

TAULUKKO 31. NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien poimimien termien tavallisimmat termikaavat

Termikaava	NaviTerm 2.0		Masterin	
	f	%	f	%
N <sub>c</sub>	49	38,0	34	29,3
N	29	22,5	21	18,1
N <sub>(c)</sub> N <sub>(c)</sub>	14	10,9	16	13,8
A N <sub>(c)</sub>	11	8,5	10	8,6
Abb N, Abb-N	10	7,7	15	12,9
Abb (en)	10	7,7	8	6,9
en	5	3,9	11	9,5
Yhteensä	128	100	115	100

NaviTerm 2.0 -terminpoimintaohjelma pisteyttää termiehdokkaat portaattomasti asteikolla 0–1, niin että paras termiehdokas saa pistemäärän 1,0. Ohjelman tunnistamista termeistä korkeimman pistemäärän 1,0 sai lyhenne *VTS* ja lähes yhtä korkeaan pistemäärään 0,942 ylsi lyhenne *VTMIS*. Nämä lyhenteet ovat ymmärrettävästi saaneet korkeat termiyspisteet, koska niillä on kymmeniä esiintymiä suhteellisen lyhyessä tekstissä. Sen sijaan yhden esiintymän lyhennetermi *SAR* on saanut valituista termeistä alhaisimmat pisteet. Lyhenteet eivät siis automaattisesti saa korkeita termiyspisteitä.

Seuraavien kahdeksan korkeimmat pisteet saaneen termiehdokkaan joukossa on seitsemän yksisanaista termiehdokasta, joiden frekvenssi lähtötekstissä on 6–14. Korkeat termiyspisteet saaneista yksisanaisista termiehdokkaista kaksi on yhdyssanoja: *laivaliikenne* ja *meriliikenne*. Muut tämän ryhmän termiehdokkaat, *satama*, *liikenne*, *turvallisuus*, *laiva* ja *tehokkuus*, ovat käsittehierarkiassa korkealla olevien käsitteiden nimityksiä. Yllättävää on, että *alus*, jonka frekvenssi lähtötekstissä on viisi, on saanut varsin alhaisen pistemäärän 0,1579. Myös termit *ohjaus*, *onnettomuus*, *operaattori*, *sisääntuloväylä* ja *tutka-asema* ovat saaneet alle 0,3 pistettä, vaikka niiden frekvenssi on vähintään viisi. Kolmen esiintymän termin *lasti* termiys on arvioitu vain pisteen sadasosien arvoiseksi. Kymmenen korkeimmat termiyspisteet saaneen termiehdokkaan joukossa on vain yksi sanaliittotermi, *meriliikenteen ohjausjärjestelmä*, joka on saanut pistemäärän 0,518, koska sillä on neljä esiintymää lähtötekstissä.

Suurifrekvenssiset yksisanaiset termiehdokkaat saavat ohjelman poimimista termiehdokkaista korkeimmat termiyspisteet, kun taas vain yhden esiintymän yksisanaiset termiehdokkaat saavat kauttaaltaan alle 0,1 termiyspistettä. Monisanaiset termiehdokkaat saavat yli 0,1 pistettä, vaikka niiden frekvenssi olisi alhainen ( $f = 1$ ). Termiyspisteet korreloivat siis positiivisesti frekvenssin ja sanaluvun kanssa. Termiehdokaslistassa on kuitenkin muutama alhaiset termiyspisteet saanut suurifrekvenssinen termi.

NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa on 92 termiä niistä 118 termistä, jotka suomalaiset asiantuntijat valitsivat heti yksimielisesti termilistaansa. Näistä 92 termistä 47 eli joka toinen asiantuntijoiden mielestä kiistaton termi on saanut alle 0,1 pistettä. NaviTerm 2.0 -ohjelman antamat termiyspisteet eivät siis ole kovin hyvä keino parantaa poimintatulosta. Termiehdokaslistaan tulostuvien termiehdokkaiden määrän voi puolittaa 340:een jättämällä kaikki vähemmän kuin 0,1 pistettä saaneet termiehdokkaat pois listalta. Tällöin kuitenkin menetetään noin puolet termeistä. Jos pisteraja asetetaan 0,2:een, listaan tulostuu 135 termiehdokasta eli vain 17 enemmän kuin asiantuntijoiden heti valitsemien termiehdokkaiden listassa, mutta saanti jää 29 termiin. Tämäkin tulos on ehkä riittävä, jos tavoitteena on poimia tekstistä vain kaikkein keskeisimmät termit. Vähintään 0,2 pistettä saaneiden termiehdokkaiden listassa ovat

esimerkiksi lyhenteet *VTS*, *VTMIS*, *IMO* ja *IALA*, substantiivitermit *satama*, *liikenne*, *turvallisuus* ja *laiva*, yhdyssanasubstantiivit *laivaliikenne*, *meriliikenne* ja *merikuljetus* sekä sanaliittotermit *meriliikenteen ohjausjärjestelmä* ja *liikenteen informaatiopalvelu*. Poimintatuloksessa ei ole mitään oleellista eroa, vaikka tarkastelussa olisivat kaikki asiantuntijoiden yhteisesti valitsevat 189 termiä, sillä edelleen noin kolmasosa termeistä on saanut yli 0,2 pistettä ja noin puolet alle 0,1 pistettä.

Olen edellä tarkastellut NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien poimimien termien ominaisuuksia. Seuraavaksi selvitän, millaisia ovat termit, joita ohjelmat eivät tunnistanee termeiksi. Piiloterminien sanalukua, frekvenssiä ja termikaavoja tarkastelemalla pyrin selvittämään, onko niillä tunnusomaisia piirteitä, joiden avulla ali-poiminta on selitettävissä.

#### 6.7.3.2 NaviTerm 2.0 ja Masterin-ohjelmien piiloterminien ominaisuuksia

NaviTerm 2.0 -terminpoimintaohjelman tuottamasta termiehdokaslistasta puuttuu 59 suomalaisten asiantuntijoiden termilistaansa valitsemaa termiä. Näistä 13 on isokirjainlyhenteen ja substantiivin yhdistelmiä, jotka on tekstissä kirjoitettu ilman yhdysmerkkiä, kuten esimerkiksi *VTS operaattori*. Tästä syystä ohjelma tunnisti termeiksi erikseen isokirjainlyhenteen ja seuraavan substantiivin, mutta ei niiden yhdistelmää.

NaviTerm 2.0 -ohjelman tuottamasta termiehdokaslistasta puuttuvien termien sanaluku on keskimäärin 2,1 sanaa ja suurin osa puuttuvista termeistä on yhden tai kahden sanan termiä. Puolet kaksisanaisista piilotermeistä on lyhenteen ja substantiivin yhdistelmiä, jotka on lähtötekstissä kirjoitettu ilman terminosia yhdistävää yhdysmerkkiä. Neljässä suomenkielisessä tunnistamatta jääneessä termissä on neljä sanaa, joista yksi on konjunktio *ja*. Termit *laivaliikenteen ohjaus- ja informaatiojärjestelmä*, *laivaliikenteen ohjaus- ja informaatiopalvelu*, *meriliikenteen ohjaus- ja informaatiopalvelu* sekä *sataman johto- ja informaatiojärjestelmä* jäivät poimimatta, koska NaviTerm 2.0 -terminpoimintaohjelma pyrkii tunnistamaan 1–3 sanan substantiivilausekkeita. Sanaliittoterminien osat *informaatiojärjestelmä* ja *informaatiopalvelu* ovat kuitenkin ohjelman termiehdokaslistassa termeinä. Tunnistamatta jäivät myös sanaliittotermit *integroitu tutkaverkko*, *integroitu verkko*, *maalle rakennettava tutka-asema*, *maalle rakennettu tutka-asema* ja *meritse tapahtuva kuljetus*, joissa on määritteenä verbin partisipiipimuoto. Suurin osa puuttuvista termeistä on ohjelman termiehdokaslistassa joko pidemmässä tai lyhyemmässä muodossa kuin asiantuntijoiden valitsema termi. Pisimmät puuttuvat termit ovat englanninkieliset termit *Vessel Traffic Management and Information Service* ja *Vessel Traffic Management and Information System*.

Masterin-ohjelman tuottamasta termiehdokaslistasta puuttuu eniten yksisanaisia termejä. Vaikka Masterin-ohjelman saanti on kaiken kaikkiaan noin seitsemän prosenttia alhaisempi kuin NaviTerm 2.0 -ohjelman, vähintään kaksisanaisista termeistä sen saanti on korkeampi kuin NaviTerm-ohjelman. Vähintään kaksisanaisista termeistä ohjelmalta jäi poimimatta 26, kun NaviTerm-ohjelmalta niistä jäi poimimatta 41.

TAULUKKO 32. NaviTerm 2.0 ja Masterin-ohjelmien piilotermien pituus

Sanaluku	NaviTerm 2.0 (A=2,10)		Masterin (A=1,53)	
	f	%	f	%
1	18	30,5	47	64,4
2	27	45,8	17	23,3
3	8	13,5	5	6,8
≥ 4	6	10,2	4	5,5
Yhteensä	59	100	73	100

NaviTerm 2.0 -terminpoimintaohjelman tuottaman termiehdokaslistan piilotermien frekvenssi on keskimäärin 1,02, eli tämän ryhmän termit esiintyvät lähtötekstissä keskimäärin kerran. Piilotermien listassa on peräti 15 nollafrekvenssistä yksisanaista termiä (*etsintä, etsintäviranomainen, hinaus, kuljetusketju, lastaus, luotsipalvelu, ohjailla, pelastus, pelastusviranomainen, päällystö, rannikkovartiosto, raportointimuoto, satamaviranomainen, sensori, turvallisuusriski*). Tietokoneohjelmalle on tietysti mahdoton tehtävä tunnistaa merkkijono, jota lähtötekstissä ei edes ole. Jos saanti lasketaan olettaen, että termiehdokaslistasta puuttuisivat täysin vain nämä 15 termiä, päästään prosenttilukuun 92,1, jota voidaan pitää erinomaisena tuloksena. Tämä on mahdollista saavuttaa pienillä manuaalisilla termiehdokaslistan korjaustoimenpiteillä. Parempaan tulokseen ei tietääkseni ole päästy millään puoliautomaattisella termintunnistusohjelmalla, eikä se ehkä ole mahdollistakaan.

Olen aiemmin todennut, että tekstissä on monta lyhenteen ja substantiivin muodostamaa termiä, joiden poiminta tuottaa vaikeuksia terminpoimintaohjelmalle. Tähän ryhmään kuuluvista termeistä yhdeksän esiintyy lähtötekstissä vain kerran. Kaksi lyhenteen ja substantiivin muodostamaa termiä esiintyy tekstissä vähintään kolme kertaa, mutta koska nekin on kirjoitettu ilman yhdysmerkkiä, ne eivät ole NaviTerm 2.0 -ohjelman termiehdokaslistassa. Yhteistä yli kaksi kertaa esiintyville puuttuville termeille on joko poikkeava kirjoitusasu tai harvinainen termikaava.

Masterin-terminpoimintaohjelman termiehdokaslistasta puuttuvat termit eivät juuri eroa frekvenssiltään NaviTerm-ohjelman termiehdokaslistasta puuttuvista termeistä. Selkein ero on yksifrekvenssisten piilotermien poiminnassa, joita Masterin-ohjelman termiehdokaslistasta puuttuu kymmenen enemmän kuin NaviTerm-ohjelman listasta.

TAULUKKO 33. NaviTerm 2.0 ja Masterin-ohjelmien piilotermien frekvenssi

Frekvenssi	NaviTerm 2.0 (A=1,02)		Masterin (A=1,08)	
	f	%	f	%
0	15	25,4	17	23,3
1	35	59,3	45	61,6
2	7	11,9	8	11,0
≥ 3	2	3,4	3	4,1
Yhteensä	59	100	73	100

NaviTerm 2.0 -terminpoimintaohjelman tuottamasta termiehdokaslistasta puuttuvien termien yleisin termikaava on lyhenteen ja substantiivin yhdistelmä, joka on tekstissä kirjoitettu ilman yhdysmerkkiä. Tällaisia termejä on puuttuvien listalla 13. Lisäksi listalla on lyhenteen tai paremmin kahden lyhenteen ja substantiivin yhdistelmä *EEZ/LRT*-



*järjestelmä*, jossa terminosat on yhdistetty yhdysmerkillä. Muuten ohjelman tuottamasta listasta puuttuvat termit eivät eroa termiehdokaslstaan tulostuneista tai asiantuntijoiden valitsemista termeistä mitenkään erityisesti. Toiseksi eniten puuttuu yhdyssubstantiiveja ja kolmanneksi eniten kahden substantiivin sanaliittotermejä. Yhden sanan substantiiveja puuttuu kuusi, samoin kuin englanninkielisiä termejä. Adjektiivin ja substantiivin sanaliittotermejä listasta puuttuu viisi. Lisäksi listasta puuttuu verbitermi *ohjailla*.

Ero NaviTerm 2.0 ja Masterin-ohjelmien välillä ilmenee selvästi piilotermin termikaavoissa. NaviTerm-ohjelmalle tunnistamisvaikeuksia tuottaneet ilman yhdysmerkkiä kirjoitetut lyhenteen ja substantiivin muodostamat termit ja englanninkieliset termit eivät Masterin-ohjelmalle olleet ongelma. Sen sijaan ohjelmalla oli vaikeuksia yhdyssubstantiivien ja joidenkin yhden sanan substantiivien tunnistamisessa. Suurin osa näistä on pienifrekvenssisia ( $f \leq 1$ ), mutta joukossa on myös kahdeksan kertaa lähtötekstissä esiintyvä termi *ohjaus* ja neljän esiintymän *ohjausjärjestelmä*.

TAULUKKO 34. NaviTerm 2.0 ja Masterin-ohjelmien piilotermin tavallisimmat termikaavat

Termikaava	NaviTerm 2.0		Masterin	
	f	%	f	%
Abb N, Abb-N	14	23,7	9	12,3
N <sub>c</sub>	10	16,9	25	34,3
N <sub>(c)</sub> N <sub>(c)</sub>	8	13,6	7	9,6
N	6	10,2	14	19,2
en	6	10,2	0	0,0
A N <sub>(c)</sub>	5	8,5	6	8,2
Yhteensä	49	83,1	61	83,6

Olen edellä tarkastellut NaviTerm 2.0 ja Masterin-terminpoimintaohjelmien tuottamista termiehdokaslstoista puuttuvien termien sanalukua, frekvenssiä ja termikaavoja ja todennut, että NaviTerm 2.0 -ohjelmalla on vaikeuksia poimia pienifrekvenssisia ( $f \leq 1$ ) termejä, joissa on useimmiten kaksi sanaa, ja termejä, joilla on poikkeava kirjoitusasu. Masterin-ohjelmalla on vaikeuksia erityisesti pienifrekvenssisten yhdyssubstantiivien poiminnassa. Seuraavaksi tarkastelen ohjelmien poimimia valetermejä.

### 6.7.3.3 NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien ominaisuuksia

NaviTerm 2.0 -terminpoimintaohjelman tuottamassa termiehdokaslistassa on 565 valetermiä. Nämä ovat sanaluvultaan keskimäärin 1,59 sanan mittaisia eli liki yhtä pitkiä kuin asiantuntijoiden poimimat termit, joissa on keskimäärin 1,54 sanaa. Ohjelma poimii liikaa varsinkin yhden ja kahden sanan termiehdokkaita, mutta liki 600 valetermin listaan on päätyntä myös 82 vähintään kolmen sanan termiehdokasta.

Masterin-ohjelma taas tuottaa 562 valetermiä, kun termiehdokaslistasta poistetaan ne valetermit, jotka esiintyvät useammassa kuin yhdessä taivutusmuodossa. Valetermien määrä on tällöin miltei sama kuin NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa. Listat näyttävät nopeasti silmäillen hyvin samanlaisilta, mutta täysin samoja termiehdokkaita niissä on kuitenkin vain 361 eli noin 64 prosenttia. Masterin-ohjelman poimimista valetermeistä suurin osa eli lähes puolet on kahden sanan mittaisia ja vain noin kolmannes on yksisanaisia.

TAULUKKO 35. NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien pituus

Sanaluku	NaviTerm 2.0 (A=1,59)		Masterin (A=1,85)	
	f	%	f	%
1	324	57,3	213	37,9
2	159	28,1	254	45,2
3	71	12,6	71	12,6
≥ 4	11	2,0	24	4,3
Yhteensä	565	100	562	100

Ohjelmien poimimat yksisanaiset valetermit ovat suurimmaksi osaksi yleiskielen sanoja, kuten sanat *aika*, *ihminen*, *kehitys* ja *ongelma*, mutta joukossa on myös terminosia ilman määriteosaa, kuten *informaatiojärjestelmä*, *suojaus* ja *verkko*. Vähintään kaksisanaiset valetermit, kuten *ihmisen mielikuvitus*, ovat päässeet termiehdokaslistalle, koska niiden syntaktinen rakenne sopii johonkin ohjelman algoritmin mukaiseen termikaavaan. Yllättävää on, että yhdenkään vähintään kaksisanaisen valetermin frekvenssi ei ole kahta suurempi ja vain kymmenen (NaviTerm) tai viidentoista (Masterin) vähintään kaksisanaisen valetermin frekvenssi on kaksi.

Muutoinkaan ylipoimitujen termiehdokkaiden keskimääräinen frekvenssi ei ole erityisen korkea. NaviTerm-ohjelman valetermeillä se on 1,48 eli lähes yhtä esiintymää pienempi kuin suomalaisten asiantuntijoiden valitsemien termien keskimääräinen frekvenssi 2,34. Ohjelmat tuottavatkin runsaasti yhden esiintymän valetermejä.

TAULUKKO 36. NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien frekvenssi

Frekvenssi	NaviTerm 2.0 (A=1,48)		Masterin (A=1,62)	
	f	%	f	%
1	456	80,7	462	82,2
2	60	10,6	52	9,3
3	17	3,0	12	2,1
≥ 4	32	5,7	36	6,4
Yhteensä	565	100	562	100

Puolet NaviTerm-ohjelman tuottamista pienifrekvenssisistä valetermeistä on yleiskielisiä yksisanaisia substantiiveja tai yhdyssubstantiiveja. Kolmannes on kaksisanaisia kahden substantiivin tai adjektiivin ja substantiivin sanaliittoja, jotka eivät termeille tyypillisestä syntaktisesta rakenteestaan huolimatta ole termejä, kuten sanaliitot *nykyinen tilanne*, *suora yhteys*, *ongelmien laajuus* ja *sisällön koostumus*. Jäljelle jäävät 79 yksifrekvenssisistä valetermiä ovat 3–5 sanan sanaliittoja, joissa terminosana on usein lyhenne tai konjunktio *ja*, kuten valetermeissä *EU:n jäsenvaltioiden edustaja* sekä *viittaus ja vertaus*. Myös kaksifrekvenssisistä valetermeistä suurin osa on yksisanaisia substantiiveja tai yhdyssubstantiiveja. Liki kaikki suurifrekvenssiset ( $f \geq 3$ ) valetermit ovat yleiskielisiä yhden kantasanan substantiiveja, kuten *alue*, *järjestelmä*, *kehitys* ja *tieto*.

NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa tavallisimmat termikaavat ovat lähes samat kuin asiantuntijoiden valitsemien termien kaavat (ks. jakso 6.3.1). Suurin ero on *ja*-konjunktin sisältävien termiehdokkaiden huomattavasti suurempi

osuus valetermien listassa. Termikaavaan sopiva termiehdokas ei kuitenkaan välttämättä ole termi. NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa yliopimittujen termiehdokkaiden tavallisimmat termikaavat ovat substantiivi tai yhdyssubstantiivi, joita listassa on 47,1 prosenttia (*aika, edistysaskel*), genetiivimuotoisella substantiivilla alkavat sanaliitot (*asiantuntijoiden yleinen mielipide*), adjektiivialkuiset sanaliitot (*asianomainen viranomainen, selvä yhteistyötoimikunta*), sanaliitot, joissa kahta substantiivia yhdistää *ja*-rinnastuskonjunktio (*järjestelmä ja tietokanta*), lyhenteet tai lyhennealkuiset termiehdokkaat (*EU-maa*) ja englanninkieliset termiehdokkaat (*Industries Forum*). Yksilökäsitteiden nimityksiksi luokiteltavia termiehdokkaita listassa on 29 (*Englannin kanaali*).

Masterin-ohjelman tuottamien valetermien yleisimmät termikaavat eivät eroa NaviTerm-ohjelman termikaavoista, sillä suurin osa ohjelman valitsemista valetermeistä on yksisanaisia substantiiveja tai yhdyssubstantiiveja. Myös kaksisanaisia kahden substantiivin tai adjektiivin ja substantiivin sanaliittoja on runsaasti. Mutta eroakin on. Masterin-ohjelma poimi kymmeniä valetermejä, jotka alkavat lukusanalla tai pronominilla (*1950-luku, kahden tutkasensori, nämä toimija*) sekä vähintään kolmisanaisia sanaliittoja, joiden osana on lyhenne (*edistysaskeleena VTS toiminta*). Näiden termikaavojen mukaiset termiehdokkaat eivät asiantuntijoiden terminvalinnan perusteella ole koskaan termejä. Masterin-ohjelman termiehdokkaista niitä on yhteensä liki sata.

TAULUKKO 37. NaviTerm 2.0 ja Masterin-ohjelmien poimimien valetermien tavallisimmat termikaavat

Termikaava	NaviTerm 2.0 f	%	Masterin f	%
N	177	31,3	128	22,8
N <sub>c</sub>	89	15,8	67	11,9
N <sub>(c)</sub> N <sub>(c)</sub>	58	10,3	82	14,6
A N <sub>c</sub>	55	9,7	70	12,5
en	30	5,3	6	1,1
Abb X	30	5,3	43	7,6
N <sub>(c)</sub> C N <sub>(c)</sub>	26	4,6	0	0,0
Yhteensä	450	82,3	396	70,5

Valetermien listasta helposti tunnistettavia ovat numeraalialkuiset termiehdokkaat (*1950-luku*), erisnimet ja englanninkieliset termiehdokkaat. Lyhennealkuiset termit on käytävä läpi manuaalisesti, koska asiantuntijoiden termilistassa on runsaasti lyhennealkuisia termejä ja lyhennetermejä. Koetekstissä niitä on ehkä jopa poikkeuksellisen paljon.

Yliopimittujen termiehdokkaiden listassa olevat sanaliitot, joissa kahta substantiivia yhdistää *ja*-sana, eivät tämän kokeen perusteella ole termejä. Suomen kielessä yhdysmerkki ja konjunktio *ja* paljastaa ellipsin. *Ja*-sana saattaa kuulua termiin, mutta siinä tapauksessa sitä edeltävän sanan lopussa on lähes aina yhdysmerkki, vaikka poikkeuksiakin on, esimerkiksi *etsintä ja pelastus*. Konkordanssihaku ratkaisisi sellaisten termiehdokkaiden termiyden, joissa on konjunktio *ja*. Jos konjunktio *ja* liittyy aina yhteen saman sanaparin, se todennäköisesti kuuluu termiin.

Yliopimittujen termiehdokkaiden listassa ongelmallisimpia ovat yksisanaiset substantiivitermiehdokkaat, adjektiivialkuiset sanaliittotermiehdokkaat ja genetiivimuotoisella substantiivilla alkavat sanaliittotermiehdokkaat, joiden syntaktinen rakenne sopii

yleisimpiin termikaavoihin. Mitkä yksisanaista termiehdokkaista ovat tarkasteltavan erikoisalan termejä ja mitkä yleiskielen sanoja? Yksisanaisten substantiivitermiehdokaiden tunnistamisessa ei termiehdokkaiden frekvenssistä ole apua, sillä jotkin termit esiintyvät tekstissä vain kerran ja moni yleiskielen sana esiintyy tekstissä monta kertaa. Lahtisen mukaan (sähköpostitse 1.3.2004) NaviTerm 2.0 -ohjelman algoritmi sisältää suuresta korpuksista ennalta lasketun geneerisen painokertoimen IDF. Sen tarkoitus on erottaa termit yleiskielen sanoista, mutta ohjelman tuottama valettermien lista osoittaa, että termiehdokkaiden tunnistamisessa tarvitaan silti erikoisalan tuntemusta. (Vrt. Lahtinen 2000: 12, 174, 182.) Genetiivimuotoisella substantiivilla alkavien valettermien määrää voisi pienentää hyväksymällä tästä ryhmästä termeiksi vain sanaliitot, joiden pääsana on emotermin eli termin yksinäänkin. Adjektiivialkuisien sanaliittotermiehdokkaiden määriteosina on esimerkiksi adjektiiveja *asianomainen*, *määräaikainen*, *uusi*, *alueellinen*, *hyvä*, *kansainvälinen*, *kansallinen*, *oma* ja *paikallinen*. Roskaluokkaan jäävät adjektiivialkuiset sanaliittotermiehdokkaat esiintyvät muutamaa poikkeusta lukuun ottamatta (*asianomainen viranomainen*, *kansallinen taso*, *uusi teknologia*) tekstissä vain kerran. Silloin kun adjektiivimäärite kuuluu termiin, sanaliitto on todennäköisesti hyvin pysyvä. Tätä hypoteesia tulisi kuitenkin testata laajemmalla aineistolla, sillä suurin osa asiantuntijoiden termeiksi hyväksymistä adjektiivialkuisista sanaliittotermeistäkin esiintyy koetekstissä vain kerran.

Sanaluvun, frekvenssin ja termikaavojen lisäksi tarkastelin NaviTerm 2.0 -ohjelman valettermeille antamia termiyspisteitä. Termiehdokaslistan valettermit ovat saaneet keskimäärin 0,1058 pistettä. Termien keskimääräinen pisteluku on 0,1625, joten valettermien termiys on keskimäärin pienempi kuin termien, niin kuin pitääkin olla. Yli puolet eli 51,7 prosenttia valettermeistä onkin saanut alle 0,1 termiyspistettä ja vain 16,5 prosenttia on saanut yli 0,2 termiyspistettä. Yli 0,5 termiyspistettä ovat saaneet lyhenteet *EU* ja *EY* sekä substantiivit *palvelu*, *yhteistyö*, *koulutus* ja *tutkimus*, joilla kaikilla frekvenssi on vähintään viisi. Vastaavat käsitteet ovat toki tärkeitä merenkulun turvallisuuden kannalta, mutta ne eivät ole merenkulun turvallisuuskäsitteitä.

Seuraavaksi tutkin, miten saanti muuttuu, jos tarkkuutta parannetaan jättämällä alle 0,1 termiyspistettä saaneet termiehdokkaat vertailun ulkopuolelle. Ohjelman tuottamaan 694 termiehdokkaan listaan jää näin 340 termiehdokasta eli noin puolet alkuperäisestä määrästä. Listaa voisi vielä lyhentää 55 termiehdokkaalla jättämällä erisnimiksi ja indeksitermeiksi luokitellut termiehdokkaat vertailun ulkopuolelle. En kuitenkaan tehnyt niin, koska listassa on 28 englanninkielistä termiehdokasta tai englanninkielistä lyhennettä, jotka ohjelma on luokitellut erisnimiksi. Korkeimmat termiyspisteet saaneiden 340 termiehdokkaan joukossa on 67 sellaista termiä, jotka asiantuntijat valitsivat termeiksi. Saanti on tällöin 36 prosenttia ja tarkkuus 19,7 prosenttia. Roskan määrä on siis edelleen hyvin suuri ja saanti on pudonnut dramaattisesti 68,3 prosentista 36 prosenttiin. Jos tavoite on saavuttaa mahdollisimman hyvä saanti, on ohjelmalla tuotettava pitkä termiehdokaslista ja käytävä se läpi manuaalisesti.

#### 6.7.4 MultiTerm Extract -ohjelman tuottamien termiehdokaslistojen ja asiantuntijalistojen vertailu

MultiTerm Extract -terminpoimintaohjelma tuotti valituilla asetuksilla suomenkielisestä artikkelista *pw02* termiehdokaslistan, jossa on 306 termiehdokasta. Kun listassa olleet saman kantasanan eri taivutusmuodot palautettiin perusmuotoon manuaalisesti, jäi listaan 255 termiehdokasta. Näistä 48 on myös asiantuntijoiden termilistassa. Asiantuntijoiden valitsemista termeistä ohjelmalta jäi poimimatta 141 termiä. Kun ohjelman poimimien termien lukumäärä 48 jaetaan asiantuntijoiden poimimien termien lukumäärällä

189, saadaan saanniksi 25,4 prosenttia. Ohjelma tunnisti siis noin joka neljännen termin. Ohjelman poimintatuloksen tarkkuudeksi saadaan 18,8 prosenttia, jos tunnistettujen termien määrää verrataan perusmuotoon palautettujen termiehdokkaiden määrään (48:255). Yli 4/5 ohjelman poimimista termiehdokkaista on roskaa. Jos vertailulistana käytetään kaikkien termiehdokkaiden listaa, tarkkuudeksi saadaan 15,7 prosenttia ja roskan osuudeksi noin 5/6.

Venäjänkielisestä tekstistä *Eskos* MultiTerm Extract -terminpoimintaohjelma poimi 365 termiehdokasta. Käytin poiminnassa samoja asetuksia kuin suomenkielisten termiehdokkaiden poiminnassa eli muutin perusasetuksia vain säätämällä roskan määrän 75 prosenttiin. Venäläisten termiehdokkaiden listassa on vähemmän saman kantasanan eri taivutusmuotoja kuin suomalaisten termiehdokkaiden listassa, sillä vain 16 termiehdokasta on listassa kahdessa eri taivutusmuodossa. Esimerkiksi substantiivitermi *судно* on termiehdokaslistassa muodoissa *судах* ja *судов* ja sanaliittotermin *морской транспорт* muodoissa *морском транспорте* ja *морскому транспорту*. Tyypillistä on, että kumpikaan termi ei ole perusmuodossa, vaan siinä muodossa, jossa se useimmin esiintyy tekstissä. Kun nämä yhdistetään, jää listaan 349 termiehdokasta, joista 29 on venäläisten asiantuntijoiden termilistassa. Venäläisten termien saanti on 27,1 prosenttia eli pari prosenttia korkeampi kuin suomalaisten termien saanti. Venäläisten termien poiminnan tarkkuus on 8,3 prosenttia eli vieläkin huonompi kuin suomalaisten termien poiminnan tarkkuus.

Seuraavaksi tarkastelen, millainen sanaluku, frekvenssi ja termikaava on niillä termeillä, jotka MultiTerm Extract -ohjelma tunnisti, millaiset termit jäivät tunnistamatta ja millaisia ovat ohjelman valitsevat valettermit.

#### 6.7.4.1 MultiTerm Extract -ohjelman poimimien suomalaisten termien ominaisuuksia

MultiTerm Extract -ohjelman suomalaisten termiehdokkaiden listassa on 255 perusmuotoon palautettua termiehdokasta, joista 48 on suomalaisten asiantuntijoiden termilistassa. Ohjelman poimimien termien keskimääräinen sanaluku on 1,58, mikä on hyvin lähellä asiantuntijoiden valitsemien termien keskimääräistä sanalukua 1,54. Tunnistetuista termeistä 29 eli yli puolet on yksisanaisia ja 14 eli liki kolmannes on kaksisanaisia. Vähintään kolmisanaisia termejä on vain viisi ja niistäkin neljä on englanninkielisiä. Asiantuntijoiden valitsemista 14:stä vähintään kolmisanaista suomenkielisestä termistä ohjelman termiehdokaslistalla on vain *maalle rakennettu tutka-asema*.

Termin pituus ei näyttänyt vaikuttavan MultiTerm Extract -ohjelman saantiin, sillä suomalaisten asiantuntijoiden valitsemista yksisanaista termeistä ohjelma tunnisti 25 prosenttia, kaksisanaista 25,9 prosenttia, kolmisanaista 25 ja vähintään nelisanaista 28,6 prosenttia.

Kaikki MultiTerm Extract -ohjelman tunnistamat suomalaiset termit kahta lukuun ottamatta esiintyvät koetekstissä vähintään kaksi kertaa. Tunnistettujen termien keskimääräinen frekvenssi 5,02 on varsin korkea, sillä asiantuntijoiden valitsemien termien keskimääräinen frekvenssi on vain 2,34. Termillä onkin oltava varsin korkea frekvenssi, jotta se varmasti pääsee ohjelman termiehdokaslistaan. Ohjelma poimii lähes erehtymättä kaikki termit, joiden frekvenssi on vähintään kolme, ja suurimman osan termeistä, joiden frekvenssi on vähintään kaksi. Vähintään kolmifrekvenssisten termien saanti on korkea, noin 83 prosenttia, mutta jostain syystä esimerkiksi lyhennetermi VTS jää ohjelmalta poimimatta, vaikka sen frekvenssi lähtötekstissä on 49. Sen sijaan korkeintaan yhden kerran lähtötekstissä esiintyvistä termeistä ohjelman tuottamassa termiehdokaslistassa on vain yksi termi, joka sekin on englanninkielinen: *Maritime Industries Forum*.

Suurin osa MultiTerm Extract -ohjelman tunnistamista suomalaisista termeistä on substantiiveja (11) tai yhdyssubstantiiveja (13). Suomenkielisiä sanaliittotermejä on 10, esimerkiksi *liikenteen ohjaus*, ja englanninkielisen lyhenteen ja suomenkielisen sanan muodostamia termejä neljä. Ne on kirjoitettu yhdysmerkin kanssa (Abb-N) tai ilman sitä (Abb N), kuten esimerkiksi *VTS-verkko* ja *VTS komitea*. Tunnistettujen 48 termin listassa on 10 englanninkielistä termiä, joista neljä on lyhenteitä (*Off-Shore VTS*, *IALA*).

MultiTerm Extract -ohjelman tunnistamista kaksisanaista termeistä kuusi on kahden substantiivin sanaliittoja ja kolme adjektiivin ja substantiivin sanaliittoja. Suurin osa ohjelman valitsemista 89 kaksisanaista termiehdokkaasta on kielelliseltä rakenteeltaan mahdollisia termejä, mutta tästä huolimatta kaksisanaisten termien saanti on vain 16,7 prosenttia. Koska ohjelma poimii termiehdokkaita frekvenssin perusteella, jää suuri osa asiantuntijoiden valitsemista kaksisanaista termeistä termiehdokaslistan ulkopuolelle liian alhaisen frekvenssin takia. Asiantuntijoiden valitsemista kaksisanaista termeistä vain 16 eli vajaa kolmannes esiintyy lähtötekstissä vähintään kaksi kertaa.

Ohjelman tunnistamat 48 termiä ovat saaneet 60–99 termiyspistettä. Keskimääräinen pistemäärä on 75, ja kahdeksan termiä on saanut maksimipisteet eli 99 termiyspistettä. Nämä ovat termit *laiva*, *laivaliikenteen ohjaus*, *meriliikenteen ohjausjärjestelmä*, *satama*, *tietojen vaihto*, *Vessel Traffic Management and Information Service*, *VTMIS* ja *VTS toiminta*. Korkeat termiyspisteet korreloivat positiivisesti termiiden kanssa, sillä suomalaiset asiantuntijat valitsivat nämä termit yksimielisesti heti ensimmäisellä kierroksella englanninkielistä termiä *Vessel Traffic Management and Information Service* lukuun ottamatta. Muut keskiarvopisteiden yli yltäneet termit ovat *IALA*, *IMO*, *maalle rakennettu tutka-asema*, *meriliikenne*, *operaattori*, *sisääntuloväylä*, *tehokkuus*, *turvallisuus*, *tutka-asema* ja *VTS komitea*. Näistä *maalle rakennettu tutka-asema* ja *tehokkuus* eivät ole asiantuntijoiden heti yksimielisesti valitsemien termien joukossa, mutta ovat kuitenkin asiantuntijoiden neuvottelemassa vertailulistassa. Kaikki ohjelman tunnistamat termit eivät kuitenkaan ole saaneet korkeita termiyspisteitä, sillä yli puolet eli 30 termiä on saanut alle 75 pistettä. 60 termiyspistettä ovat saaneet esimerkiksi kuuden esiintymän *laivaliikenne* ja viiden esiintymän *viranomainen*.

#### 6.7.4.2 MultiTerm Extract -ohjelman suomalaisten piilotermien ominaisuuksia

MultiTerm Extract -ohjelmalta jäi poimimatta 141 termiä suomalaisten asiantuntijoiden valitsemasta 189 termistä eli alipoiminta on 74,6 prosenttia. Poimimatta jääneet termit eivät keskimääräiseltä sanaluvultaan eroa poimituista termeistä. Piilotermien keskimääräinen sanaluku on 1,53, poimitujen termien keskimääräinen sanaluku on 1,58 ja asiantuntijoiden poimimien termien keskimääräinen sanaluku on 1,54. Alipoiminta kohdistuikin tasaisesti yksi- ja monisanaisiin termeihin. Eripituisten termien ryhmissä puuttuvien termien osuus on 74–80 prosenttia sanaluvusta riippumatta. Ainoa poikkeus on kuusisanaisten termien ryhmä, jossa puuttuvien termien osuus on 50 prosenttia, sillä ohjelmalta jäi poimimatta toinen kahdesta kuusisanaista englanninkielisestä termistä. Piilotermien osuus on pienin kaksisanaisten termien ryhmässä (74,1 %) ja suurin nelisanaisten termien ryhmässä (80 %).

Eripituisten piilotermien lukumäärien tarkastelu kuvaa alipoimintaa selkeämmin kuin prosenttilukujen tarkastelu. Yksi-, kaksi- ja kolmisanaista termeistä ohjelma jättää poimimatta peräti kolme neljäsosaa. Nelisanaista termeistä jää poimimatta vieläkin suurempi osuus eli neljä viidestä. Ohjelma kyllä poimii eripituisia termiehdokkaita, sillä esimerkiksi yksisanaista vaalestermejä on termiehdokaslistassa 189 eli yhtä paljon kuin asiantuntijoiden valitsemia termejä. Toisaalta 75 prosenttia yksisanaista termeistä jää



ohjelmalta poimimatta, sillä yksisanaisten termien sijasta ohjelma poimii yleiskielen sanoja.

Silmiinpistävä piirre piilotermien listassa on näiden termien hyvin pieni frekvenssi, tavallisesti yksi. Ohjelma ei käytännöllisesti katsoen pystykään tunnistamaan alle kaksifrekvenssisia termejä, sillä ohjelma tunnistaa vain yhden 122 tällaisesta termistä.

Saanti nousee frekvenssin kasvaessa, niin että kaksifrekvenssisistä termeistä ohjelmalta jää poimimatta 14 ja kolmifrekvenssisistä vain kolme termiä. Niistä 23 termistä, joiden frekvenssi on vähintään neljä, jää poimimatta ainoastaan kolme. Suuri frekvenssi ei kuitenkaan aina riitä takaamaan termiehdokaslistalle pääsyä, sillä ohjelman termiehdokaslistasta puuttuva lyhennetermi *VTS* esiintyy koetekstissä 49 kertaa.

MultiTerm Extract -ohjelman valitsemien termiehdokkaiden termikaavat poikkeavat selvästi asiantuntijoiden valitsemien termien kaavoista. Ohjelma ei ole poiminut juuri ollenkaan pitkiä sanaliittotermejä, kuten voi olettaa termiehdokkaiden sanaluvun perusteella (vrt. Pasanen 2005b). Tavallisimpienkin termikaavojen mukaisista termeistä jää ohjelmalta poimimatta suuri osa. Esimerkiksi kaavan  $A N_{(c)}$  mukaisista 16 termistä jää poimimatta 13 ja kaavan  $N_{(c)} N_{(c)}$  mukaisista 23 termistä 17. Epätavallisten termikaavojen mukaisia termejä ei ohjelma poiminut ollenkaan. Tällaisia ovat esimerkiksi termit, joissa on konjunktio *ja* (esimerkiksi *laivaliikenteen ohjaus- ja informaatiopalvelu*).

Puolet piilotermistä on yhdyssubstantiiveja tai muita substantiiveja, joita asiantuntijoiden termilistassa on 94. Näistä 70 jää poimimatta. Ohjelma ei myöskään pystynyt tunnistamaan termeiksi yhdeksää ilman yhdysmerkkiä ja kymmentä yhdysmerkin kanssa yhteen kirjoitettua isokirjainlyhenteen ja substantiivin yhdistelmää. Kaikkiaan asiantuntijoiden listassa on 23 tällaista termiä, joten ohjelma poimi niistä vain joka kuudennen. Myös lyhennetermit ovat ohjelmalle ongelmallisia, sillä asiantuntijoiden valitsemista 10 englanninkielisestä isokirjainlyhenteestä poimimatta jää kuusi. Noin viidennes piilotermistä on kahden sanan sanaliittoja, joissa jälkimmäisenä osana on yhdyssubstantiivi tai muu substantiivi ja määritteenä on adjektiivi tai toinen substantiivi.

Manuaalisesta poiminnasta ja muista tässä tutkimuksessa tarkastelluista ohjelmista poiketen MultiTerm Extract -ohjelma poimii termiehdokkaiksi runsaasti myös muita kuin substantiiveja tai substantiivilausekkeita. Termiehdokaslistassa on muun muassa useita verbejä. Tästä huolimatta sitä ainoaa verbiä (*ohjailla*), jonka asiantuntijat valitsivat termiksi, ei termiehdokaslistassa ole.

#### 6.7.4.3 MultiTerm Extract -ohjelman poimimien suomalaisten valetermien ominaisuuksia

MultiTerm Extract -ohjelman tuottamassa manuaalisesti perusmuotoon palautettujen 255 termiehdokkaan listassa on 207 valetermiä. Niistä ylivoimaisesti suurin osa eli yli 90 prosenttia on yhden sanan pituisia. Keskimääräinen sanaluku on 1,14, eli merkittävästi pienempi kuin poimittujen termien sanaluku 1,58 ja piilotermien sanaluku 1,53. Lisäksi ohjelma poimii 2–4-sanaisia valetermejä, joiden frekvenssi lähtötekstissä on kaksi.

Ohjelma tuottaa runsaasti suurifrekvenssisia valetermejä, sillä valetermien keskimääräinen frekvenssi on 3,56. Pienifrekvenssiset sanat tai sanaliitot eivät yleensä päädy valetermien listaan, sillä valetermeistä vain 14 on yksifrekvenssisia. Yli 40 prosenttia valetermeistä onkin termiehdokkaita, joiden frekvenssi lähtötekstissä on kaksi, ja yli puolet valetermeistä esiintyy lähtötekstissä vähintään kolme kertaa. Listassa on myös 47 sellaista termiehdokasta, jonka frekvenssi on vähintään viisi.

Yhden kerran lähtötekstissä esiintyvät valetermit ovat yksisanaisia yleiskielisiä substantiiveja, verbejä tai verbin partisiippimuotoja. Kaksi kertaa esiintyvät ovat 1–4-sanaisia, mutta termiehdokkaan tai sen osien sanaluokat eivät noudata suomalaisten

termien yleisimpiä termikaavoja. Tämän ryhmän valetermien joukossa on runsaasti adjektiiveja, adverbeja ja verbejä. Vähintään kolme kertaa esiintyvät valetermit ovat yhtä poikkeusta lukuun ottamatta yksisanaisia yleiskielen sanoja, joiden joukossa on runsaasti adjektiiveja ja verbejä.

MultiTerm Extract -terminpoimintaohjelman poimimien valetermien termikaavat poikkeavat selvästi asiantuntijoiden poimimien termien termikaavoista. Vaikka suuri osa eli 36,2 prosenttia ylipoimituista termiehdokkaista on substantiiveja tai yhdyssubstantiiveja, samoin kuin asiantuntijoidenkin poimimista termeistä, puolet valetermeistä on adjektiiveja (21,3 %), verbejä (16,4 %) tai adverbeja (12,1 %). Asiantuntijoiden termilistassa sen sijaan on vain yksi verbi ja adjektiivejakin vain termin osana.

Valetermit ovat saaneet termiyspisteitä 60–99, ja keskimääräinen pistemäärä on 68 eli hieman pienempi kuin tunnistetuilla termeillä, joilla keskimääräinen pistemäärä on 75. Vaikka valetermit ovat yleensä saaneet alhaisia termiyspisteitä, niiden joukossa on 22 vähintään 90 pistettä saanutta termiehdokasta ja 48 yli 75 pistettä saanutta termiehdokasta. Korkeimmat pisteet (99) ovat saaneet esimerkiksi valetermit *järjestelmä*, *kansallisella tai kansainvälisellä tasolla*, *kuitenkin*, *luku*, *määritelmä*, *oleva*, *palvelu*, *palvelujen laajuutta kuvaava käsite*, *sama*, *tietty*, *tulla* ja *voida*. Ohjelman antamat termiyspisteet eivät siis ole hyvä termiuden mittari, koska kaikki korkeita pisteitä saaneet termiehdokkaat eivät ole termejä.

MultiTerm Extract -ohjelma on tarkoitettu lähinnä teknisten termien poimintaan, kun taas NaviTerm 2.0 -ohjelma on tarkoitettu indeksitermien, muun muassa erisnimien, tunnistamiseen. Tämä ero ilmenee poiminnan tuloksista, sillä MultiTerm Extract -ohjelman poimiman roskan joukossa on vähemmän erisnimiä kuin NaviTerm 2.0 -ohjelman tuottamassa termiehdokaslistassa eikä ollenkaan numeroilla alkavia termiehdokkaita.

#### 6.7.4.4 MultiTerm Extract -ohjelman poimimien venäläisten termien ominaisuuksia

Venäjänkielisestä lähtötekstistä *Eskos* MultiTerm Extract -terminpoimintaohjelma poimi 365 termiehdokasta. Näistä 16 termiehdokkaassa on samat termielementit, mutta pääasana on kahdessa eri taivutusmuodossa. Kun nämä termiehdokkaat yhdistetään, listaan jää 349 termiehdokasta, joista 29 on asiantuntijoiden valitsemia termejä. Poimitujen termien keskimääräinen sanaluku on 2,10 eli hieman alhaisempi kuin venäläisten asiantuntijoiden poimimien termien keskimääräinen sanaluku 2,36. Yksisanaisten termien osuus on 37,9 prosenttia, kun se asiantuntijoiden termilistassa on 26,1 prosenttia. Asiantuntijoiden valitsemista 28 yksisanaista termistä ohjelma tunnistikin 11 eli noin 40 prosenttia, mutta kaksi- ja kolmisanaista termeistä ohjelma tunnisti vain joka neljännen. Termin pituuden kasvaessa saanti putoaa vieläkin alhaisemmaksi, sillä vähintään nelisanaisten poiminnassa saanti jää 18,8 prosenttiin eli tunnistamatta jää noin neljä viidesosaa vähintään nelisanaista termeistä. Suomenkielisellä aineistolla tehty havainto, että MultiTerm Extract -ohjelma tunnistaa monisanaista termejä huomattavasti enemmän kuin yksisanaista, pätee siis myös venäläiseen tekstiin.

MultiTerm Extract -ohjelman venäjänkielisestä koetekstistä tunnistamat termit esiintyvät tekstissä keskimäärin 7,86 kertaa. Keskimääräistä frekvenssiä kasvattavat 12 keskeistä termiä, joilla on tekstissä yli viisi esiintymää. Suurin osa tunnistetuista termeistä esiintyy tekstissä vähintään kaksi kertaa. Ohjelmalla on vaikeuksia erityisesti yksi- ja kaksifrekvenssisten termien tunnistuksessa, sillä asiantuntijoiden valitsemista 56 yksifrekvenssisestä termistä ohjelman termiehdokaslistalla on vain neljä ja 14 kaksifrekvenssisestä termistä kolme. Ohjelman poimintatulos paranee jonkin verran, jos termin

frekvenssi on vähintään kolme, sillä suurifrekvenssisistä termeistä ohjelma tunnistaa yli puolet.

MultiTerm Extract -ohjelman tunnistamien venäläisten termien termikaavat noudattavat asiantuntijoiden valitsemien termien termikaavoja, vaikka adjektiivin ja substantiivin sanaliittojen osuus, 17,2 prosenttia, on pienempi kuin asiantuntijoiden termilistassa ja substantiivien osuus, 24,1 prosenttia, huomattavasti suurempi. Ohjelma on tunnistanut termeiksi substantiivilausekkeita ja lyhenteitä, mutta myös harvinaisempia termikaavoja. Tunnistettujen termien listassa on yksi termi, jossa on konjunktio *и* (*единая система контроля и управления судоходством*) ja kaksi termiä, joissa on prepositio (*надзор за судоходством* ja *контроль над судоходством*). Yhteen tunnistettuun substantiivitermiin liittyy muu merkki kuin kirjain (*район AI*). Kaikkien termikaavaltaan harvinaisempien termien frekvenssi koetekstissä on vähintään kolme. Tunnistettujen venäläisten termien listassa on myös englanninkielinen lyhenne *VTMIS*.

MultiTerm Extract -ohjelman poimimat termit ovat saaneet 48–99 termiyspistettä. Keskimääräinen pistemäärä on 82, ja kahdeksan termiä on saanut maksimipisteet eli 99 termiyspistettä. Nämä ovat termit *ЕСКУС*, *единая система контроля и управления судоходством*, *порт*, *центр морской спутниковой связи*, *система судовых сообщений*, *служба капитана порта*, *система управления движением судов* ja *район AI*, joista venäläiset asiantuntijat valitsivat yksimielisesti heti ensimmäisellä kierroksella vain termit *порт* ja *система управления движением судов*. Yli kolmannes eli 11 termiä on saanut alle 82 pistettä. Esimerkiksi termi *судоходство*, jolla on lähtötekstissä 17 esiintymää, on saanut vain 48 termiyspistettä ja viiden esiintymän lyhenne *VTMIS* on saanut 57 pistettä. Korkeat termiyspisteet eivät siis ole luotettava termiuden osoitin.

#### 6.7.4.5 MultiTerm Extract -ohjelman venäläisten piilotermin ominaisuuksia

MultiTerm Extract -ohjelmalta jäi tunnistamatta 89 venäläistä termiä asiantuntijoiden valitsemista 107 termistä eli alipoiminta on 83,2 prosenttia. Piilotermit ovat keskimäärin 2,45 sanan mittaisia, joten alipoiminta kohdistuu erityisesti sanaliittotermeihin, joita asiantuntijoiden termilistassa on noin 74 prosenttia. Asiantuntijoiden valitsemista sanaliittotermeistä MultiTerm Extract -ohjelman termiehdokaslistasta puuttuukin 77,2 prosenttia. Piiloterminien osuus on suurin vähintään nelisanaisten termien ryhmässä, sillä näistä 81,3 prosenttia puuttuu ohjelman termiehdokaslistasta. Tulos on samansuuntainen kuin suomalaisten termien poiminnassa, jossa ohjelmalta jäi poimimatta 74–80 prosenttia termeistä sanaluvusta riippumatta. Venäläisten piiloterminien osuus on alhaisin yksisanaisissa termeissä. Silti ohjelman tuottamasta termiehdokaslistasta puuttuu 60,1 prosenttia yksisanaisistakin termeistä.

Suomenkielisellä aineistolla tehdyn poimintakokeen perusteella oli odotettavissa, että myös venäläisten piiloterminien frekvenssi on hyvin pieni. Näin onkin, sillä 57,3 prosenttia piilotermeistä esiintyy lähtötekstissä vain kerran. Lähes kaikki asiantuntijoiden valitsemat pienifrekvenssiset ( $f \leq 1$ ) termit jäivät ohjelmalta poimimatta. Kaksi esiintymää ei vielä näytä takaavan termin pääsyä ohjelman termiehdokaslistalle, mutta frekvenssin kasvaessa mahdollisuudetkin kasvavat. Tästä huolimatta piiloterminien listassa ovat esimerkiksi suurifrekvenssiset lyhenteet *ССС*, *АИС* ja *СУДС*.

MultiTerm Extract -ohjelmalta jää poimimatta suurin osa tavallisimpienkin termikaavojen mukaisista venäläisistä termeistä. Esimerkiksi kaavan A N mukaisista 22 termistä jää poimimatta 17 ja kaavan N mukaisista 16 termistä yhdeksän. Substantiivialkuisista sanaliittotermeistä jää poimimatta 26 eli 70,3 prosenttia asiantuntijoiden termilistan 37 substantiivialkuisesta sanaliittotermistä.

Ohjelmalla on suuria vaikeuksia myös adjektiivialkuisten termien poiminnassa, sillä termiehdokaslistasta puuttuu 28 asiantuntijoiden valitsemista 32 adjektiivialkuisesta termistä. Tämä selittyy näiden termien alhaisella frekvenssillä. Sen sijaan yllättävää on, että termiehdokaslistasta puuttuu kahdeksan lyhennetermiä, joista kuuden frekvenssi on vähintään kaksi. Lyhennetermit ovat siis ohjelmalle ongelmallisia, kuten myös suomalaisten termien poimintatulokset osoitti (ks. jakso 6.7.4.2)

#### 6.7.4.6 MultiTerm Extract -ohjelman poimimien venäläisten valetermien ominaisuuksia

MultiTerm Extract -ohjelman tuottamassa manuaalisesti perusmuotoon palautettujen 349 venäläisen termiehdokkaan listassa on 320 valetermiä. Keskimääräinen sanaluku on 1,68 eli merkittävästi pienempi kuin venäläisten asiantuntijoiden poimimien termien sanaluku 2,36. Suurin osa eli noin 60 prosenttia valettermeistä on yhden sanan pituisia. Kolmannes valettermeistä on kaksi- tai kolmisanaisia, mutta listassa on myös 15 nelisanaista ja kuusi vähintään viisisanaista valetermiä. Lähes kaikkien vähintään nelisanaisten valetermien frekvenssi on vähintään kaksi.

Venäläisten valetermien frekvenssi on keskimäärin 4,31 eli korkeampi kuin venäläisten termien. Ohjelma tuottaa runsaasti suurifrekvenssisia ( $f \geq 4$ ) yksisanaisten termiehdokkaita, jotka kuuluvat yleiskieleen (esimerkiksi *информация, система, данные*), mutta eniten eli 44,4 prosenttia valetermien listassa on kaksifrekvenssisia termiehdokkaita, joista suurin osa on pitkiä, jopa yhdeksän sanan sanaliittoja. Joka neljän valetermien frekvenssi on kuitenkin vähintään viisi. Yksifrekvenssisia valettermeistä on vain 26 eli alle 10 prosenttia.

MultiTerm Extract -terminpoimintaohjelman poimimien valetermien termikaavat noudattavat asiantuntijoiden poimimien termien kaavoja, vaikka joidenkin termikaavojen mukaisia termiehdokkaita on suhteettoman paljon. Selkein ero on adjektiivien runsaus valetermien listassa, vaikka asiantuntijoiden termilistasta adjektiivi yksinään ei ole koskaan termi. Termiehdokaslistaan ovat päätyneet esimerkiksi suurifrekvenssiset adjektiivit *бассейный, морской, региональный* ja *технический*. Toinen selkeä ero on yksisanaisten substantiivien suuri osuus valettermeistä, joista niitä on lähes kolmannes. Suurin osa yksisanaisten valettermeistä on suurifrekvenssisia yleiskielen substantiiveja, kuten esimerkiksi *информация, контроль, обеспечение* ja *центр*. Kaksisanaisten valetermien ryhmässä on adjektiivin ja substantiivin tai kahden substantiivin sanaliittoja, jotka voisivat olla termejä kielelliseltä rakenteeltaan, mutta eivät kuitenkaan ole merenkulun turvallisuuden termejä, kuten *правовая основа* (oikeudellinen perusta), *региональная система* (alueellinen järjestelmä), *система связи* (viestintäjärjestelmä) ja *точка зрения* (näkökulma).

Venäläisten asiantuntijoiden termilistassa on eniten adjektiivin ja substantiivin muodostamia sanaliittotermejä. MultiTerm Extract -ohjelman valetermien listassa on kaavan A N mukaisia termiehdokkaita, mutta niin on myös piilotermien listassa. Sama pätee kahden substantiivin sanaliittoihin. Ohjelma poimii siis termikaavoihin sopivia termiehdokkaita, mutta valitettavasti väärinä. Selittävä tekijä on jälleen frekvenssi, sillä ohjelma poimii sanaliittoja, joiden frekvenssi on vähintään kaksi, kun taas asiantuntijoiden poimimien termien frekvenssi on yleensä vain yksi.

Aiemmin on jo todettu, että lyhenteet ovat ongelmallisia MultiTerm Extract -ohjelmalle. Se poimii kyllä lyhenteitä, mutta valitettavasti valitsee väärät lyhenteet. Tällä kertaa frekvenssi ei riitä selittäväksi tekijäksi, sillä vain kahden puuttuvan lyhenteen frekvenssi on yksi. Ohjelma poimii myös runsaasti substantiivin ja lyhenteen muodostamia sanaliittoja, vaikka niitä on asiantuntijoiden termilistassa vain yksi (*радиомаяк DGPS*). Kolmisanaisten valettermeistä lyhenne on termin osana jopa joka kolmannessa

ja samoin joka kolmannessa termin osana on konjunktio *u* tai *или*. Valetermien listaan päätyneistä termiehdokkaista 23:ssa termi tai termin osa on sanaluokaltaan verbi, yhdeksässä adverbi ja kolmessa pronomini. Näihin sanaluokkiin kuuluvia sanoja ei asiantuntijoiden valitsemien termien listassa esiinny lainkaan.

Valetermit ovat saaneet 48–99 termiyspistettä eli skaalan alimmasta pistemäärästä skaalan ylimpään, ja keskimääräinen pistemäärä on 68 eli hieman pienempi kuin tunnistetuilla termeillä, joilla keskimääräinen pistemäärä on 82. MultiTerm Extract -ohjelman pisteytys on varsin suoraviivainen, sillä termiyspisteet nousevat frekvenssin ja sanaluvun noustessa. Kaikki pitkät, suurifrekvenssiset valetermit ovat saaneet erittäin hyvät termiyspisteet, joko 98 tai 99 pistettä. Tämän ryhmän termiehdokkaiden termikaavat poikkeavat muiden termiehdokkaiden termikaavoista siten, että termin osana on usein prepositio, pronomini tai verbi.

Alhaisimmat pisteet saaneet valetermit ovat kaikki yksisanaisia. Yksisanaisten valetermien pistemäärä on kuitenkin suoraan verrannollinen frekvenssiin, niin että viisifrekvenssiset yksisanaiset valetermit saavat 82 termiyspistettä. Korkeimmat termiyspisteet ohjelma antaa valetermeille, joilla on poikkeuksellisen korkea frekvenssi tai joiden pituus on vähintään kolme sanaa.

Olen edellä verrannut kielittäin ja ryhmittäin koehenkilöryhmien ja puoliautomaattisten terminpoinintaohjelmien tuottamia termiehdokkaita asiantuntijoiden valitsemiin termeihin. Vertailussa olen tarkastellut termien ja termiehdokkaiden pituutta sanoina, frekvenssiä koetekstissä ja termikaavoja. Seuraavaksi esitän tämän vertailun tulokset kootusti ja arvioin, minkälaisia tuloksia erilaisilla termintunnistusmenetelmillä saadaan ja miten tuloksia voi hyödyntää.

## **6.8 Terminpoinintamenetelmien vertailu ja arviointi**

Luvussa 5 pohdin termin olemusta terminpoininnan näkökulmasta ja totesin, että termin määritelmät eivät anna riittävästi keinoja termien erottamiseksi muista sanoista. Tutkimuskirjallisuudessa esitetyt ihannetermin vaatimukset eivät nekään auta tässä tehtävässä, koska ne ovat enemmänkin toivottuja ominaisuuksia, jotka tulee ottaa huomioon silloin, kun nimetään uusia käsitteitä. Jo olemassa olevien termien tunnistuksessa niistä ei ole juuri apua. Termeillä ei myöskään ole mitään muotopiirteitä, joiden avulla ne voisi erottaa muista sanoista, sillä sanojen tai sanaliittojen termiys perustuu ennen kaikkea niiden semanttiseen funktioon erikoiskielessä eikä muotoon tai kieliopilliseen rakenteeseen. Tätä taustaa vasten on ymmärrettävää, että erikoisalan asiantuntijoidenkin terminvalinnassa voi olla suuria eroja, kuten tähän tutkimukseen kuulunut terminpoininnan pilottikoe osoitti (ks. jakso 6.3). Terminpoininnan tulosten arviointi riippuu siis olennaisesti siitä termilistasta, johon eri menetelmillä tuotettuja termiehdokaslistoja verrataan. Tässä tutkimuksessa termiehdokaslistoja on verrattu asiantuntijoiden koostamiin termilistoihin luottaen siihen, että asiantuntijoiden terminvalinta on niin luotettava ja perusteltu kuin terminvalinta ylipäätään voi olla.

Eri terminpoinintamenetelmillä tuotettujen termiehdokaslistojen saanti ja tarkkuus eroavat huomattavasti. Seuraavaksi arvioin termiehdokkaiden pituuden, frekvenssin, termikaavan ja variaation vaikutusta eri menetelmillä tehdyn terminpoininnan saantiin ja tarkkuuteen.



### 6.8.1 Saanti ja tarkkuus terminpoiminnassa

Eräässä aiemmassa tutkimuksessa on todettu, että kielenpuhujat tunnistavat termit helposti tekstistä eikä tulos riipu edes erilaisista koulutustaustoista (Kobrin 1989: 23, 27). Koehenkilöiden terminpoiminta ei perustu muodollisiin tai syntaktisiin termin ominaisuuksiin, vaan koehenkilöt poimivat termejä käsitelähtöisesti. Kobrinin tutkimustuloksen perusteella koehenkilöiden tekemän terminpoiminnan tulos on todennäköisesti hyvin lähellä asiantuntijoiden terminpoimintaa. Oletettavasti saanti ja tarkkuus on korkea kaikissa koehenkilöryhmissä.

Manuaalisessa terminpoiminnassa saanti ja tarkkuus vastasivat pääpiirteittäin ennakko-oletuksia. Suomalaisten merenkulun opiskelijoiden valitsemien termiehdokkaiden listassa on 128 sellaista termiä, jotka myös asiantuntijat valitsivat. Saanti tässä koehenkilöryhmässä on siis 67,7 prosenttia ( $128 : 189 \times 100$ ). Suomalaisten kääntämisen opiskelijoiden valitsemien termiehdokkaiden listassa on 137 asiantuntijoiden valitsemaa termiä, jolloin saanti on 72,5 prosenttia ( $137 : 189 \times 100$ ). Merenkulun opiskelijoiden yhdistetyssä termiehdokaslistassa on 63 ja kääntämisen opiskelijoiden termiehdokaslistassa 101 valetermiä eli sellaista termiehdokasta, joita asiantuntijat eivät valinneet termeiksi. Merenkulun opiskelijoiden terminvalinnan tarkkuudeksi saadaan 67,0 ( $128 : 191 \times 100$ ) ja kääntämisen opiskelijoiden 57,6 ( $137 : 238 \times 100$ ) prosenttia. Merenkulun opiskelijoiden termiehdokaslistassa on siis toisaalta vähemmän roskaa kuin kääntämisen opiskelijoiden termiehdokaslistassa, mutta toisaalta saanti on alhaisempi kuin kääntämisen opiskelijoiden ryhmässä. Termien tunnistamiseen liittyvä ero ryhmien välillä oli melkein merkitsevä (Chin neliö -testi,  $p=0,027344$ ).

Venäläisten kieliaineiden opiskelijoiden termiehdokaslistassa on 82 täysin samaa termiä kuin asiantuntijoiden valitsemien termien listassa. Tässä ryhmässä saanti on 76,6 prosenttia ( $82 : 107 \times 100$ ). Venäläisten merenkulun opiskelijoiden valitsemien termiehdokkaiden listassa on 74 sellaista termiä, jotka myös venäläiset merenkulun tai terminologian asiantuntijat valitsivat. Merenkulun opiskelijoiden ryhmässä saanti on 69,2 prosenttia ( $74 : 107 \times 100$ ). Merenkulun opiskelijoiden poiminnan saantia voi pitää varsin hyvänä. Ryhmän poimimien 253 termiehdokkaan joukossa on kuitenkin 179 valetermiä, jolloin poiminnan tarkkuus jää 29,2 prosenttiin ( $74 : 253 \times 100$ ). Kieliaineiden opiskelijoiden poimimien termiehdokkaiden joukossa valetermejä on 79 eli huomattavasti vähemmän kuin merenkulun opiskelijoiden termiehdokaslistassa. Tässä ryhmässä tarkkuus nouseekin 50,9 prosenttiin ( $82 : 161 \times 100$ ). Joka toinen ryhmän poimima termiehdokas on todella termi. Termien tunnistamiseen liittyvä ero ryhmien välillä oli erittäin merkitsevä (Chin neliö -testi,  $p=4,58E-05$ ).

Sekä venäläisistä että suomalaisista koehenkilöryhmistä korkeampaan saantiin ylti kieliaineiden opiskelijoiden ryhmä. Ryhmien väliset erot ovat kuitenkin vain muutamia prosenttia ja saanti on kaikilla koehenkilöryhmillä vähän alle tai vähän yli 70 prosenttia. Poiminnan tarkkuudessa erot ryhmien välillä ovat suuremmat, venäläisten koehenkilöryhmien välillä eroa on noin 20 prosenttia. Venäläisten merenkulun opiskelijoiden poiminnan tarkkuus on varsin alhainen. Tulos on yllättävä, sillä suomalaisilla koehenkilöillä poiminnan tarkkuus on parempi merenkulun opiskelijoiden ryhmässä. Terminpoiminnassa terminologian teorian tuntemuksesta näyttäisi siis olevan enemmän hyötyä kuin erikoisalan tuntemuksesta, ainakin jos tavoitteena on korkea saanti. Yllättävää on, että koehenkilöryhmien termiehdokaslistoista puuttuu vähintäänkin noin neljännes asiantuntijoiden valitsemista termeistä. Piilotermit ovat yleensä vakiintumattomia erikoisalan käsitettä kuvaavia sanayhdistelmiä, joiden kielellinen variaatio on runsasta (ks. variaatiosta tarkemmin jakso 6.8.5).



Puoliautomaattinen terminpoiminta tietokoneohjelmilla perustuu yksinomaan tilastollisiin, muodollisiin tai syntaktisiin kielen yksiköiden ominaisuuksiin. Vaikka semanttisia ominaisuuksia ei oteta huomioon, on terminpoimintaohjelmien saanti noussut aiemmissa tutkimuksissa jopa yli 90 prosenttiin (Soininen 1999b: 9). Esimerkiksi Soinisen (1998: 7) raportoimassa terminpoimintakokeessa testattu ohjelma löysi englanninkielisestä lähtötekstistä lähes kaikki toistuvat monisanaiset termiehdokkaat. L’Hommen ym. (1996: 307) kokeessa vain 3,8 prosenttia termeistä jäi poimimatta, kun aineistona oli lennonjohtoa käsitteleviä tekstejä. Neurologian teksteistä jäi poimimatta 11,7 prosenttia termeistä. Aina ei saanti kuitenkaan ole ollut yhtä hyvä, sillä esimerkiksi Laakson (2003: 50) kokeessa se oli 36,5 prosenttia, eli testattu ohjelma poimi vain noin kolmasosan termeistä. Terminpoimintakokeessani testatuista ohjelmista ei mikään yltänyt lähellekään 90 prosentin saantia. NaviTerm 2.0 -termintunnistusohjelman saanti on 68,3 eli jopa hieman korkeampi kuin suomalaisten merenkulun opiskelijoiden saanti, mutta tarkkuus on kovin alhainen. Masterin-ohjelman saanti 61,4 prosenttia on lähes yhtä hyvä, sillä sen tuottama termiehdokaslista eroaa vain vähän NaviTerm-ohjelman listasta. MultiTerm Extract -terminpoimintaohjelman saanti on alhaisin, vain 25,4 prosenttia suomalaisista ja 27,1 prosenttia venäläisistä termeistä.

Tietokoneohjelmalla tehdyn terminpoiminnan tarkkuus on aiemmissa englanninkielisillä aineistoilla tehdyissä tutkimuksissa vaihdellut 30 ja 70 prosentin välillä (L’Homme ym. 1996: 303–307), mutta esimerkiksi Sewangin (2001: 88) kokeileman lingvistisen menetelmän tarkkuus oli 21 prosenttia eli 79 prosenttia termiehdokkaista oli roskaa. Aiemmin mainitussa Laakson (2003:50) kokeessa tarkkuus oli vain 15 prosenttia. Tarkkuus vaihtelee siis huomattavasti ja keskimäärin noin puolet terminpoimintaohjelmien poimimista termiehdokkaista on termejä (L’Homme ym. 1996: 303–307). Kaikilla testaamillani terminpoimintaohjelmilla poiminnan tarkkuus jäi alle 20 prosenttiin. Heikoimpaan tulokseen ylti jälleen MultiTerm-ohjelma, jonka poimimista venäjänkielisistä termiehdokkaista yli 90 prosenttia on valettermejä. Puoliautomaattisessa terminpoiminnassa tulokset vaihtelivat huomattavasti ohjelmien välillä, mutta kielten välillä vaihtelua oli yllättävän vähän. Seuraavaan taulukkoon on koottu terminpoimintakokeen saanti ja tarkkuus eri kielissä ja eri menetelmillä.

TAULUKKO 38. Terminpoiminnan saanti ja tarkkuus

Termiehdokaslista	Kieli	Termiehdokkaita	Saanti %	Tarkkuus %
kofi	fi	238	72,5	57,6
mofi	fi	191	67,7	67,0
NaviTerm 2.0	fi	694	68,3	18,6
Masterin	fi	690	61,4	16,8
MultiTerm	fi	255	25,4	18,8
koru	ru	161	76,6	50,9
moru	ru	253	69,2	29,2
MultiTerm	ru	349	27,1	8,3

## 6.8.2 Termin pituuden vaikutus terminpoimintaan

On väitetty, että suurin osa erikoiskielen termeistä on kahden sanan substantiivilausekeita (esim. Katz 1996: 25; Pazienza 1998/1999: 184; Vivaldi & Rodríguez 2001: 35). Tämän tutkimuksen tulokset kuitenkin viittaavat siihen, että merenkulun suomessa reilusti yli puolet termeistä on yksisanaisia ja vain runsas neljännes kaksisanaisia (vrt. Pasanen 2005b). Tämä selittyy kielen luonteella, sillä suomi on agglutinoiva kieli. Tämän kielen piirteen vuoksi suuri osa suomenkielisistä yksisanaista termeistä on yhdyssanoja. Vain noin 10 prosenttia termeistä on kolmen sanan pituisia tai pitempiä.

Suomalaiset koehenkilöt olivat varsin yksimielisiä suomalaisten asiantuntijoiden kanssa eripituisten termien osuudesta, vaikka merenkulun opiskelijat poimivat suhteellisesti enemmän yksisanaisia termiehdokkaita kuin kääntämisen opiskelijat. Sen sijaan puoliautomaattiset terminpoimintaohjelmat tuottivat suomalaisesta tekstistä runsaasti varsinkin yksisanaisia valetermejä. Aiemmissa tutkimuksissa on todettu, että terminpoimintaohjelmilla on vaikeuksia yksisanaisten termien poiminnassa, sillä moni ohjelma ei hae yksisanaisia termiehdokkaita (ks. esim. Soininen 1999a: 13–15). NaviTerm 2.0 -ohjelma tunnisti kuitenkin yksisanaiset termit jopa paremmin kuin koehenkilöt, mutta tuotti myös erittäin paljon yksisanaisia valetermejä. MultiTerm Extract -ohjelman tuottamasta termiehdokaslistasta sen sijaan puuttuu erittäin paljon yksisanaisia termejä, vaikka ohjelma poimii runsaasti yksisanaisia valetermejä. Jos terminpoimintaohjelma ei tunnista yksisanaisia termejä, saanti on auttamatta alhainen, sillä asiantuntijoiden terminlistan mukaan valtaosa suomenkielisistä termeistä on yksisanaisia. Ratkaisuksi on esitetty yksisanaisten termien frekvenssin vertaamista yleiskielen korpukseseen (Urizar ym. 2000: 374). Tämä on kuitenkin mahdollista vain, jos terminpoimijalla on käytössään yleiskielen korpus.

Kumpikin suomalainen koehenkilöryhmä poimi lähes kaikki vähintään nelisanaiset termit ja yli puolet kolmisanaista termeistä. Pitkien termien saanti nousi siis korkeaksi, mutta tarkkuus jäi alhaiseksi, koska koehenkilöt poimivat pitkiä valetermejä suunnilleen saman verran kuin pitkiä termejäkin. Puoliautomaattisilla terminpoimintaohjelmilla oli vaikeuksia kaksisanaisten ja sitä pitempien suomalaisten termien poiminnassa. Näistä termeistä puuttuu kaikkien ohjelmien tuottamista termiehdokaslistoista enemmän kuin niitä on listoissa. NaviTerm 2.0 -terminpoimintaohjelman tuottamassa termiehdokaslistassa on monisanaisia termeiltä kuulostavia ja termikaavaan sopivia termiehdokkaita, kuten termiehdokkaat *seuraavan käyntisataman maan asianomainen viranomainen* ja *informaation antajan tietojen kaupallinen arvo*, joissa adjektiivin ja substantiivin sanaliittoa (asianomainen viranomainen, kaupallinen arvo) määrittää kolme peräkkäistä genetiivimuotoista määritettä. Osa termiehdokkaista on liian lyhyitä, kuten *verkko* ja *tietojenvaihto*, sillä termin statuksen ne saavat asiantuntijoiden mielestä vasta määritteiden kanssa: *tutka-asemien verkko*, *horisontaalinen tietojenvaihto* ja *vertikaalinen tietojenvaihto*. (Vrt. Laakso 2003: 50).

Venäjässä termeistä on tutkimukseni perusteella yksisanaisia vain noin neljännes, kun taas kaksisanaisia ja vähintään kolmisanaisia termejä on kumpiakin yhtä paljon eli reilu kolmannes. Venäläiset koehenkilöryhmät valitsivat melkein kaikki yksisanaiset termit, mutta monisanaisten termien saanti pieneni termin pituuden kasvaessa. Vähintään nelisanaisten termien ryhmässä piilotermien osuus on suurempi kuin poimittujen termien osuus. Venäläisten koehenkilöryhmien poiminta poikkesi suomalaisten opiskelijoiden valinnasta, sillä suomalaisten koehenkilöryhmien poiminnan saanti oli alhaisin yksisanaisissa termeissä mutta parani termin pituuden kasvaessa. Myös MultiTerm Extract -terminpoimintaohjelman venäläisten termien saanti on paras yksisanaisten termien

ryhmässä ja heikoin vähintään nelisanaisten termien ryhmässä. Suuri yksisanaisten valetermien määrä heikentää kuitenkin ohjelman poimintatulosta.

Tutkijat ovat esittäneet, että puoliautomaattinen terminpoiminta tulisi aloittaa pitkistä termeistä, koska liian pitkistä muodoista on helppo karsia ylimääräiset sanat pois. Myös sisäkkäisten termien haku tulisi aloittaa pitkistä termiehdokkaista (Pearson 1998: 125). Tämä menetelmä vaatii kuitenkin runsaasti manuaalista työtä. Muravitskaja ja Perebejnos (1987: 86) ratkaisisivat termin pituuden ongelman konkordanssin avulla, koska konkordanssihaun tulosteessa näkyy termiehdokkaan tyypillinen ympäristö. Nykyisellään terminpoimintaohjelmien tuottama termiehdokaslista on niin puutteellinen, että tarvittavaa manuaalista työtä helpottaisi, jos ohjelmissa olisi myös konkordanssihakumahdollisuus. Silloin liian lyhyet termiehdokkaat voisi täydentää konkordanssihaulla.

Tutkimukseni tulosten perusteella on selvää, että sanaluku ei riitä ratkaisemaan termiehdokkaan termiyyttä. Seuraavaksi tarkastelenkin terminpoimintakokeessa tuotettujen termiehdokkaiden frekvenssin ja termiuden yhteyttä.

### **6.8.3 Termin frekvenssin vaikutus terminpoimintaan**

Yleisesti on hyväksytty väite, että frekvenssiltään keskinkertaiset sanat ovat merkittävimpiä ja kantavat eniten tietoa (ks. esim. Lahtinen 2000: 74). Jos tämä väite voidaan ulottaa koskemaan myös termejä, niin sanastotyössä ja terminologisessa tutkimuksessa tulisi kiinnittää eniten huomiota frekvenssiltään keskinkertaisiin termeihin ja tutkia, minkälaisia käsitteitä ne nimeävät ja miten ne sijoittuvat käsitejärjestelmään. Puoliautomaattisten terminpoimintaohjelmien suunnittelussa onkin yleensä lähtökohtana oletus, että termit esiintyvät erikoisalan tekstissä vähintään kaksi kertaa. Tämä oletamus on kuitenkin jo joissakin aiemmissa tutkimuksissa (ks. esim. Sewangi 2001: 94) todettu vääräksi. Moni termi esiintyy tekstissä vain kerran tai termin sijasta käytetään synonyymia, varianttia tai lyhennettä (Pearson 1998: 123). Sewangin (2001: 94) vertailu tilastollisen puoliautomaattisen terminpoimintamenetelmän ja manuaalisen poiminnan välillä osoitti, että suurin osa termeistä esiintyy teksteissä vain kerran. Tutkimukseni tulokset tukevat Sewangin tutkimustulosta (vrt. Pasanen 2005b). Suomalaisten asiantuntijoiden valitsemista termeistä 65,6 prosenttia ja venäläisten asiantuntijoiden valitsemista termeistä 52,3 prosenttia esiintyy lähtötekstissä 0–1 kertaa. Tästä syystä terminpoimintaohjelmalta jää suurin osa termeistä poimimatta, jos ohjelman algoritmilla on ehto, että termiehdokkaan frekvenssin lähtötekstissä on oltava vähintään kaksi. Lähtötekstin pituus saattaa tosin vaikuttaa tulokseen. Yksifrekvenssisten termien ohittaminen on kieltämättä houkuttelevaa, sillä jos yhden esiintymän termit jätetään huomiotta, nousee esimerkiksi NaviTerm 2.0 -terminpoimintaohjelman saanti 86:een ja tarkkuus 34 prosenttiin. MultiTerm Extract -ohjelman saanti nousee 71 prosenttiin, mutta tarkkuus jää 19 prosenttiin.

Koehenkilöillä oli vaikeuksia pienifrekvenssisten termien poiminnassa, sillä nolla- ja yksifrekvenssisten termien osuus suomalaisista piilotermeistä on noin 80 prosenttia ja venäläisistä noin 65 prosenttia. Kaikki pienifrekvenssiset termit eivät kuitenkaan ole piilotermejä, sillä nolla- ja yksifrekvenssisten termien osuus suomalaisten koehenkilöiden valitsemista termeistä on noin 60 prosenttia ja venäläisten koehenkilöiden valitsemista termeistä lähes 50 prosenttia. Valestermeistäkin suurin osa on yksifrekvenssisia kummassakin kielessä ja kaikissa koehenkilöryhmissä, joten pieni frekvenssi ei ole estänyt termiehdokkaan valintaa. Suuri frekvenssi on kuitenkin manuaalisessa poiminnassa edistänyt termiehdokkaan valintaa, sillä koehenkilöiltä on jäänyt valitsematta vain pieni osa vähintään kaksifrekvenssisistä termeistä ja valestermeistä huomattava osa esiintyy lähtötekstissä vähintään neljä kertaa.

Puoliautomaattisessa poiminnassa NaviTerm 2.0 tuottaa erittäin runsaasti vaalestermejä, joiden frekvenssi lähtötekstissä on korkeintaan yksi. Jopa yli 80 prosenttia vaalestermeistä on yksifrekvenssisiä. Sen sijaan MultiTerm Extract -ohjelma poimii hyvin vähän yksifrekvenssisiä termiehdokkaita. Tästä syystä ohjelma ei tuota paljoa pienifrekvenssisiä vaalestermejä ja siltä jäävät poimimatta lähes kaikki yksifrekvenssiset suomalaiset ja venäläiset termit eli suurin osa asiantuntijoiden valitsemista termeistä. Sen sijaan ohjelma tuottaa runsaasti suurifrekvenssisiä yksisanaaisia vaalestermejä, jotka ovat paremmin yleiskielen sanoja kuin minkään erikoisalan termejä (esimerkiksi *alue, järjestelmä, kehitys, määritelmä, tieto, toiminta, yhteistyö*). Sama pyrkimys näkyy myös venäjänkielisestä lähtötekstistä tuotetussa termiehdokaslistassa. Listassa on eniten vaalestermejä, joiden frekvenssi lähtötekstissä on kaksi tai vähintään neljä.

Koska nolla- tai yksifrekvenssisten termien osuus asiantuntijoiden valitsemista termeistä on tutkimukseni perusteella yli puolet, niitä ei voi jättää termiehdokaslistan ulkopuolelle. Lisäksi harvoin esiintyvien termien konteksti saattaa sisältää enemmän terminologisesti kiinnostavaa tietoa kuin suurifrekvenssisten termien konteksti. Usein esiintyvät termit nimeävät käsitteitä, joita ei määritellä joka kerta. Sen sijaan harvoin esiintyvät termit nimeävät käsitteitä, jotka määritellään ehkä lähes jokaisen esiintymän yhteydessä. Harvoin esiintyviä termejä ei myöskään ole sanakirjoissa, jolloin niitä koskeva käsitetieto on vain teksteissä.

Pienifrekvenssisten termien poimintaan on siis keksittävä ratkaisu, joka ei tuota ylen määrin roskaa. Keinoksi ei riitä sanaluvun ja frekvenssin yhdistäminen, sillä tämä menetelmä tuottaa runsaasti yksisanaaisia yleiskielisiä vaalestermejä. Seuraavaksi tarkastelen termikaavan ja termiuden yhteyttä.

#### 6.8.4 Termikaavan vaikutus terminpoimintaan

Suomalaisilla koehenkilöillä oli vaikeuksia yksisanaisten substantiivitermien ja kaksisanaisten substantiivilausekkeiden valinnassa. Lähimmäksi asiantuntijoiden valintaa koehenkilöt pääsivät lyhennetermien, englanninkielisten termien ja lyhenteen sisältävien termien valinnassa. Merenkulun opiskelijat valitsivat yllättäen termeiksi adjektiiveja, mutta vain kolmanneksen asiantuntijoiden valitsemista adjektiivin ja substantiivin muodostamista sanaliittotermeistä. Kääntämisen opiskelijat puolestaan valitsivat kaikki termit, joissa lyhenne on terminosana ja lisäksi runsaasti lyhenteen sisältäviä vaalestermejä. Lyhenteillä onkin erikoisasema terminologiassa, sillä aiempien tutkimusten perusteella lyhenne on melko varma merkki siitä, että termi on keskeinen (esim. Soininen 1998: 13). Lyhennetermi ja termin pitkä muoto ovat tekstissä todennäköisesti hyvin lähellä toisiaan, joten termin pitkää muotoa kannattaa etsiä esimerkiksi suluissa olevan lyhenteen edeltä. Terminologiille tämä tieto on tärkeä, sillä pitkä muoto esiintyy tekstissä yleensä vain kerran jos sitäkään.

Suurin osa venäläisistä termeistä on termikaavaltaan A N, N, N N tai Abb. Näiden kaavojen mukaisista termeistä venäläiset koehenkilöt poimivat liki 90 prosenttia. Täytyy kuitenkin todeta, että myös vaalestermien määrä on suurin näiden termien ryhmässä. Venäjässä monisanaisten termien termikaavat vaihtelevat enemmän kuin suomessa, koska olen laskenut prepositiot sanalukuun. Suomessa yhdyssanat vähentävät termin sanojen lukumäärää ja samalla myös erilaisten termikaavojen määrää.

Erityisen pulmallisia niin koehenkilöille kuin terminpoimintaohjelmillekin ovat termit, joissa on rinnastuskonjunktio *ja*. Tällaisia termejä on suomalaisten asiantuntijoiden termilistassa viisi. Sekä suomalaiset koehenkilöt että terminpoimintaohjelmista varsinkin NaviTerm 2.0 poimivat termiehdokkaita, joissa terminosana on konjunktio *ja*, mutta ne eivät olleet samoja kuin asiantuntijoiden valitsemat *ja*-sanat sisältävät termit. Koe-

henkilöt poimivat esimerkiksi termiehdokkaat *ohjaus- ja informaatiopalvelu, turvallisuus- ja ympäristöriski* sekä *VTS menetelmä ja säännös* ja NaviTerm 2.0 esimerkiksi termiehdokkaat *käyttötö ja kustannus, ahtaus ja varastointi* sekä *säännös ja menetelmä*, mutta eivät termejä *etsintä ja pelastus* tai *sataman johto- ja informaatiojärjestelmä*. Tällaiset sanaliittotermit ovat terminpoimintaohjelmalle hankalia, jos niissä on enemmän kuin kolme sanaa, sillä ainakin NaviTerm 2.0 -terminpoimintaohjelmassa oletuksena on, että suomenkielisessä termissä on 1–3 sanaa.

NaviTerm 2.0 -terminpoimintaohjelma tunnistaa hyvin tavallisimpien termikaavojen mukaiset suomalaiset termit. Erittäin hyvin ohjelma onnistuu lyhenteiden poiminnassa. Ohjelman ongelma on suuri joukko valetermejä, joista suurin osa on yksi- tai kaksisanaisia yleiskielisiä substantiivilausekkeita, sillä kaikki teksteissä olevat kaavan mukaiset rakenteet eivät ole termejä. Sewangi (2001: 27) toteaaakin, että termikaavat eivät ole välttämättä vain termeille ominaisia, vaan ne edustavat pikemmin tutkittavan kielen substantiivilausekkeiden rakennetta. Lisäksi kielen monimuotoisuus heikentää rakenna-kaavojen luotettavuutta keinona termien erottamiseksi muista sanoista.

MultiTerm Extract -ohjelmalla on suuria vaikeuksia poimia minkä tahansa termikaavan mukaisia suomenkielisiä ja venäjänkielisiä termejä. Yllättäen ohjelmalla on vaikeuksia poimia venäläisiä sanaliittotermejä, jotka alkavat adjektiivimääritteellä. Ainoastaan englanninkielisten termien poiminnassa ohjelma onnistuu kohtalaisesti, sillä näistä termeistä ohjelma poimii yli puolet. Ohjelma kyllä poimii termikaavoihin sopivia termiehdokkaita, mutta valitettavasti väärinä. Selittävä tekijä on frekvenssi, sillä ohjelma poimii sanaliittoja, joiden frekvenssi on vähintään kaksi, kun taas asiantuntijoiden poimimien termien frekvenssi on yleensä vain yksi. Valetermien listassa on erittäin runsaasti yksisanaisia, suurifrekvenssisia termiehdokkaita sekä yleisimpien suomalaisten ja venäläisten termikaavojen mukaisia termiehdokkaita, mutta lisäksi valetermien listassa on muun muassa verbejä ja adverbeja sekä sanaliittoja, joissa näihin sanaluokkiin kuuluvia sanoja on terminosina, vaikka niitä ei asiantuntijoiden valitsemisessa termeissä juuri esiinny.

Terminpoimintaohjelmien soveltaman lingvistisen analyysin keinoin ei pystytäkään ratkaisemaan kaikkia terminpoiminnan ongelmia. Lingvistiset terminpoimintaohjelmat ovat voimattomia esimerkiksi silloin, kun sama sana on jonkin erikoisanalan termi ja yleiskielen sana. Kageuran ja Uminon (1996: 283) mukaan melko staattiset termin rakenteen kuvaukset voisivatkin hyötyä tutkimuksista, joissa termejä tarkastellaan teksteissä todellisessa käyttöyhteydessä. Näin termeistä muodostuisi dynaamisempi kuva. Termien tarkastelu käyttöyhteydessään eli aidoissa käyttöteksteissä vaatii paneutumaan esimerkiksi synonymiaan ja termivariaatioon, jotka ovat hyvin yleisiä, mutta riittämättömästi terminologisessa tutkimuskirjallisuudessa käsiteltyjä ilmiöitä.

### 6.8.5 Synonymia ja variaatio terminpoiminnassa

Jaksossa 5.4 totesin, että synonymia määritellään termien väliseksi suhteeksi, jossa termit nimeävät samaa käsitettä (ISO 1087-1: 2000(E/F): 8). Mutta myös termin variantit nimeävät samaa käsitettä. Synonymiaa ja variaatiota on vaikea erottaa sekä teoreettisesti että käytännössä, mutta jaksossa 5.4 esitin erotteluperusteeksi syntytapaa. Variantit ovat alkuperäisen termin muunnelmia, jotka eroavat toisistaan kirjoitusasultaan. Synonyymiset termit sen sijaan ovat leksikaalisesti erilaisia termejä, joiden synnyn taustalla on esimerkiksi näkökulmaero tai toiston välttäminen. Tutkimuksessani käsittelen synonymiaa ja variaatiota yhdessä ja pidän synonymiaa variaation yhtenä muotona. Terminpoimintakokeessa ilmenneen variaation tarkastelussa noudatan jaksossa 5.4 esiteltyä jakoa syntaktiseen, morfosyntaktiseen, semanttiseen ja ortografiseen variaatioon. Tar-



kastelen tässä jaksossa myös ellipsiä, lyhenteitä ja vieraskielisiä nimityksiä, joita voidaan pitää variaation tai synonymian ilmenemismuotoina. Seuraavaksi selvitän, poimivatko koehenkilöt ja terminpoimintaohjelmat termivariantit samalla tavalla kuin asiantuntijat.

Suomalaiset koehenkilöt poimivat termivariantteja hyvin samalla tavalla kuin asiantuntijat, eikä koehenkilöryhmienkään välillä ollut suurta eroa. Koehenkilöt tunnistivat termejä käsitelähtöisesti, sillä he poimivat yleensä joko kaikki käsitteen termivariantit tai eivät yhtään varianttia. Erityisen selvästi tämä näkyy englanninkielisten termien ja lyhennetermien valinnassa. Koehenkilöryhmät poimivat samoja käsitteitä nimeävät englanninkielisen termin, lyhennetermin ja suomenkielisen termin muutoin samalla tavoin kuin asiantuntijat, mutta merenkulun opiskelijat eivät poimineet englanninkielistä termiä *Allied Services* eivätkä sen suomenkielisiä vastineita *yhdistyneet palvelut* ja *Yhteiset Palvelut*. Terminpoimintaohjelmista NaviTerm 2.0 -ohjelman poiminta on hyvin lähellä asiantuntijoiden valintaa, sillä kuuden käsitteen erikielisistä nimityksistä ohjelma on tunnistanut asiantuntijoiden kanssa samalla tavalla kaikki muut paitsi käsitteiden *merenkulun foorumi* ja *meriliikenteen ohjaus- ja informaatiopalvelu* nimityksistä ohjelma on tunnistanut vain lyhennetermit *MIF* ja *VTMIS*. MultiTerm Extract -ohjelma tunnisti suomenkielisen nimityksen vain käsitteelle *meriliikenteen ohjausjärjestelmä*. Muiden tämän ryhmän käsitteiden nimityksistä ohjelma tunnisti englanninkielisen termin, lyhenteen tai kummankin, mutta ei suomenkielistä nimitystä.

Ortografisen, syntaktisen tai semanttisen variaation tapauksissa kääntämisen opiskelijoiden poiminta eroaa asiantuntijoiden poiminnasta neljässä tapauksessa viidestätoista siten, että kääntämisen opiskelijat poimivat vain osan asiantuntijoiden poimimista termivarianteista. Merenkulun opiskelijoiden poiminnassa tällaisia tapauksia on kolme. Opiskelijat poimivat esimerkiksi vain termin *kokonaiskuljetusketju*, mutta eivät termiä *kuljetusketju*, ja termin *informaatiopalvelu*, mutta eivät termiä *liikenteen informaatiopalvelu*. Käsitteiden *rannikkovaltio* ja *yhdistyneet palvelut* erilaisten nimitysten poiminnassa koehenkilöryhmien poiminta eroaa täysin. Kääntämisen opiskelijat poimivat termit *rannikkovaltio*, *rantavaltio*, *yhdistyneet palvelut* ja *Yhteiset Palvelut*, joita merenkulun opiskelijat eivät poimineet.

NaviTerm 2.0 -ohjelmalle ortografinen, syntaktinen tai semanttinen variaatio eivät tuottaneet suuria ongelmia. Ohjelma ei kuitenkaan tunnistanut lyhenteen ja substantiivin muodostamia termejä, jos osat on liitetty toisiinsa ilman yhdysmerkkiä, kuten termeissä *VTs operaattori* ja *VTs toiminta*. Ohjelman termiehdokaslistasta puuttuu termi *integroitu tutkaverkko* ja sen syntaktinen variantti *integroitu verkko*, mutta muulloin ohjelma on tunnistanut ainakin toisen syntaktisista varianteista. Listasta puuttuvat semanttiset termivariantit *meriliikenteen ohjaus- ja informaatiopalvelu* sekä *laivaliikenteen ohjaus- ja informaatiopalvelu*, mutta siinä ovat kaikki muut semanttisen variaation tapaukset: *rannikkovaltio* ja *rantavaltio*, *yhdistynyt palvelu* ja *Yhteinen palvelu*, *yhteistoimintakomitea*, *yhteistyökomitea* ja *yhteistyötoimikunta* sekä *lastin käsittely* ja *rahdinkäsittely*. MultiTerm Extract -ohjelmalla sen sijaan on suuria vaikeuksia varsinkin semanttisen variaation hallinnassa. Ohjelman tuottamassa termiehdokaslistassa ovat vain variantit *yhteistoimintakomitea* ja *yhteistyökomitea*. Syntaktisen ja ortografisen variaation tapauksissa termiehdokaslistassa on jokin termivariantti joka toiselle käsitteelle, mutta yhdellekään käsitteelle ei listassa ole kaikkia asiantuntijoiden valitsemia variantteja.

Suomalaiset asiantuntijat ovat valinneet termilistasta kuudelle käsitteelle kaksi eripituista nimitystä, termin pitkän muodon ja elliptisen termin. Näitä termejä koehenkilöt poimivat vaihtelevasti. Kääntämisen opiskelijat poimivat asiantuntijoiden kanssa samalla tavalla vain termiparin *päällikkö* ja *aluksen päällikkö*. Muita termejä he eivät poimineet ollenkaan tai poimivat vain joko pitkän tai lyhyen muodon. Elliptisiä ilmauksia ei



ilmeisestikään ole pidetty termivariantteina vaan aivan eri käsitteen nimityksinä. Merenkulun opiskelijat poimivat kaikki variantit tai eivät yhtään varianttia. Ainoa poikkeus on käsite *informaatiopalvelu*, joka on merenkulun opiskelijoiden listassa, mutta pitkä muoto *liikenteen informaatiopalvelu* ei.

Myös terminpoimintaohjelmien tuottamissa termiehdokaslistoissa on elliptisiä termejä vaihtelevasti. NaviTerm 2.0 -ohjelman listassa on aina vähintään yksi varianteista, mutta MultiTerm Extract -ohjelman listasta puuttuu ainakin yksi tai useampi variantti. Poikkeus on termipari *informaatiopalvelu* ja *liikenteen informaatiopalvelu*, jotka ovat MultiTerm Extract -ohjelman listassa, koska kummallakin on vähintään kaksi esiintymää.

Konjunktion *ja* sisältävien elliptisten ilmausten poiminnassa koehenkilöryhmien poiminta eroaa asiantuntijoiden poiminnasta huomattavasti. Ilmauksessa *meriliikenteen ohjaus- ja informaatiopalvelu* on neljä asiantuntijoiden poimimaa termiä: *meriliikenne*, *ohjauspalvelu*, *informaatiopalvelu* ja *meriliikenteen ohjaus- ja informaatiopalvelu*. Koehenkilöt ovat poimineet termit muuten samalla tavalla kuin asiantuntijat, mutta merenkulun opiskelijat eivät pitäneet *ohjauspalvelua* erillisenä terminä. Sen sijaan *informaatiopalvelu* on kummankin opiskelijaryhmän termiehdokaslistassa samoin kuin pitkä muoto *meriliikenteen ohjaus- ja informaatiopalvelu*. Ilmauksessa *turvallisuus- ja ympäristöriski* on asiantuntijoiden mukaan kaksi termiä, *turvallisuusriski* ja *ympäristöriski*. Kääntämisen opiskelijoiden listassa on termiehdokas *turvallisuus- ja ympäristöriski* ja merenkulun opiskelijoiden listassa vain *ympäristöriski*. *Ja*-konjunktioilla yhdistettyjen elliptisten termien poiminta tuottaa siis vaikeuksia jopa manuaalisessa poiminnassa.

Terminpoimintaohjelmille elliptisten ilmausten jakaminen termeihin on ongelma, sillä ohjelmat eivät pysty poimimaan termejä, joiden osat eivät ole tekstissä peräkkäin vaan välissä on esimerkiksi *ja*-konjunktio. Termien ja termiehdokaiden frekvenssejä käsittelevissä jaksoissa olen luokitellut tällaiset termit nollafrekvenssisiksi. Asiantuntijoiden termilistassa niitä on kaikkiaan 15. Ongelma pitäisi ratkaista, koska se on hyvin tavallinen. Ratkaisu saattaisi löytyä hyödyntämällä tietoa, että suomessa elliptisessä ilmauksessa on *ja*-sanaa edeltävän sanan lopussa usein yhdysmerkki osoittamassa poisjätettyä terminosaa.

Venäläisten asiantuntijoiden valitsemien termien listassa on vähemmän variaatiota kuin suomalaisten asiantuntijoiden termilistassa, sillä termien määräkin on huomattavasti pienempi. Venäläisen termilistan muutaman variaatiotapauksen perusteella onkin vaikea sanoa, eroaako koehenkilöryhmien poiminta oleellisesti asiantuntijoiden poiminnasta. Vähiten eroja asiantuntijoiden ja koehenkilöryhmien välillä on lyhennetermien ja termin pitkien muotojen poiminnassa. Koehenkilöryhmät poimivat yleensä kaikki samat lyhennetermit ja pitkät muodot kuin asiantuntijatkin. Merenkulun opiskelijoiden termiehdokaslistasta puuttuvat kuitenkin lyhennetermien *VTMIS* ja *CKII* pitkät termimuodot *система управления и информационного обеспечения судоходства* ja *спасательно-координационный подцентр*.

MultiTerm Extract -ohjelma poimii pikemmin pitkiä muotoja kuin lyhenteitä, jos käsitettä nimitään tekstissä sekä lyhennetermillä että termin pitkällä muodolla. Poimintatulokset on osittain selitettävissä termien frekvenssillä, mutta vain osittain. Poimittujen lyhenteiden *VTMIS*, *ЕСКУС* ja *СКЦ* frekvenssi tekstissä *Eskos* on vähintään viisi, mutta ohjelma jättää poimimatta esimerkiksi lyhenteet *СУДС* ja *ССС*, joiden frekvenssit ovat 25 ja 15. Termin pitkä muoto sen sijaan voi päästä termiehdokaslistalle, vaikka sen frekvenssi on vain yksi.

Syntaktisen variaation tapauksissa poiminnassa on vähemmän eroa opiskelijaryhmien välillä kuin asiantuntijoiden ja opiskelijoiden välillä. Ainoa ero opiskelijaryhmien välillä on termin *бассейновая информационная система* poiminnassa. Merenkulun

opiskelijat poimivat vain termin lyhyemmän muodon *бассейновая система*. Kumpikaan opiskelijaryhmä ei poiminut termejä *контроль движения судов* ja *контроль за движением судов*. Termeistä *обеспечение безопасности мореплавания* ja *обеспечение безопасного плавания* opiskelijat poimivat vain ensimmäisen. MultiTerm Extract -ohjelman termiehdokaslistassa on vain termi *обеспечение безопасности мореплавания*, jonka frekvenssi on neljä. Muiden tämän ryhmän termien frekvenssi on yksi tai kaksi. Semanttisen variaation esimerkkikäsitteelle asiantuntijat poimivat kolme nimitystä: *контроль над морским транспортом*, *контроль над судоходством* ja *надзор за судоходством*. Opiskelijat ja MultiTerm Extract -ohjelma eivät poimineet ensimmäistä termivarianttia, mutta kieliaineiden opiskelijat poimivat sanaliiton *контроль судоходства*, jota asiantuntijat eivät pitäneet terminä.

Elliptisiä termejä opiskelijaryhmät poimivat eri tavalla kuin asiantuntijat. Ainoastaan termien *морской бассейн* ja *бассейн* poiminnassa opiskelijat olivat yksimielisiä asiantuntijoiden kanssa. Termeistä *морской район А1* ja *район А1* kieliaineiden opiskelijat poimivat vain elliptisen muodon, kun taas merenkulun opiskelijat eivät pitäneet terminä kumpaakaan muotoa. Termeistä *морской флот* ja *флот* kieliaineiden opiskelijat poimivat jälleen vain elliptisen termin, mutta merenkulun opiskelijat poimivat kummankin. Suomalaisten merenkulun opiskelijoiden tavoin venäläiset merenkulun opiskelijat poimivat kaikki variantit tai eivät yhtään varianttia. MultiTerm Extract -ohjelman poiminta on suoraviivaisempi, sillä ohjelman tuottamalle termiehdokaslistalle pääsivät vain elliptiset muodot, sillä niiden frekvenssi on vähintään viisi. Tosin frekvenssi ei tässäkin tapauksessa yksin riitä selittämään poimintatulosta, sillä myös termin *морской район А1* frekvenssi on viisi.

Tutkimukseeni kuuluneen terminpoimintakokeen asiantuntijaosuuden perusteella termivariaatiota esiintyy merenkulun suomessa ja merenkulun venäjässä. Koehenkilöt valitsivat saman käsitteen erilaisia nimityksiä vaihtelevasti. Variaatio on ongelma puoliautomaattisessa terminpoiminnassa, jos termintunnistuksessa käytetään pääasiassa tilastollista termintunnistusmenetelmää. Termintunnistusoohjelmia onkin arvosteltu siitä, että niissä ei kiinnitetä huomiota variaation ongelmaan (Daille ym. 1996: 201). Taustalla saattaa vaikuttaa käsitys, että variaatio ei kuulu terminologiaan, koska ihannetilanteessa käsitteellä on vain yksi termi. Termien ajatellaan olevan ”kivettyneitä” eli pysyviä ilmaisuja, joilla ei ole muuntelua (frozen expressions without modification). Kivettyneet ilmaisut ovatkin ihanteellisia puoliautomaattisessa terminpoiminnassa, jossa termintunnistus perustuu samanlaisena toistuvien sanaliittojen frekvensseihin (vrt. Ahmad & Salway 1997: 103–104). Tähän ajatteluun liittyy kuitenkin vaara, että tekstissä ehkä kerran esiintyvät termivariantit jäävät tunnistamatta ja paljon arvokasta terminologista tietoa jää saamatta. Yksifrekvenssiset termivariantit ovat yleisiä laajassakin korpuksessa, sillä esimerkiksi Daillen ym. (1996: 198–199) kokeessa 1,5 miljoonan sanan lääketieteen korpuksesta löytyi käsitteelle *epithelial cell* 15 erilaista nimitystä, joista vain yksi esiintyi tekstissä useammin kuin kerran. Myös Jacquemin (2001: 313) tutkimus todistaa, että variaatio ei ole mitenkään poikkeuksellista. Jacquemin mielestä variaatio on monella tavalla jopa toivottavaa. Sen kautta voidaan esimerkiksi havaita ja jäljittää tieteellinen luovuus. Puoliautomaattisia terminpoimintamenetelmiä olisikin kehitettävä niin, että ne hakisivat valituille termiehdokkaille myös erilaisia variantteja.

### 6.8.6 Merenkulun turvallisuuden ydintermit terminpoiminnassa

Tutkimukseni terminpoimintakokeessa oli oletuksena, että termien joukosta erottuvat ydintermit, joiden termiyden aste on korkea ja pysyvyys suuri. Oletuksen mukaan sekä ihminen että tietokoneohjelma tunnistaa varmimmin juuri ydintermit. Jos oletus saa tukea terminpoimintakokeessa, koehenkilöt ovat valinneet ne termeiksi täysin tai ainakin lähes yksimielisesti. Oletettavasti termintunnistusohjelmat ovat antaneet näille termiehdokkaille korkeat termiyspisteet. Jos tällainen ydintermien joukko on olemassa, termintunnistusohjelmien tulisi tunnistaa ainakin nämä termit. Ydintermien määrän voi laskea karkeasti tekstin sanaluvun perusteella, sillä erään arvion mukaan (Pritchard 2003: 9) tietyn erikoisalan termien määrä on 5–7 prosenttia erikoisalatekstin sanastosta.

Suomalaiset asiantuntijat valitsivat heti ensimmäisellä valintakierroksella yksimielisesti 118 termiä ja venäläiset asiantuntijat 37 termiä. Pidän näitä termejä koetekstien ydintermeinä ja tarkastelen seuraavaksi, kuinka suuri osa koehenkilöistä valitsi nämä termit ja miten korkeat termiyspisteet ne saivat puoliautomaattisessa terminpoiminnassa. Vertasin asiantuntijoiden heti valitsemia termejä kustakin koehenkilöryhmästä vähintään seitsemän henkilön valitsemiin termiehdokkaisiin. Lisäksi tarkastelin heti valittujen termien saamia termiyspisteitä.

Vähintään seitsemän suomalaisen kääntämisen opiskelijan valitsemien 132 termiehdokkaan listassa on 81 sellaista termiä, jotka asiantuntijat valitsivat ensimmäisellä kierroksella. Saanti on tällöin 68,6 ja tarkkuus 61,4 prosenttia. Merenkulun opiskelijoiden termiehdokaslistassa on 67 sellaista termiä, jotka asiantuntijat ja vähintään seitsemän opiskelijaa valitsi heti. Tässä ryhmässä saanti on 56,8 ja tarkkuus 66,3 prosenttia. Kummassakaan ryhmässä ydintermien saanti tai tarkkuus ei nouse kovin korkeaksi. Huomattavin ero kääntämisen opiskelijoiden ja merenkulun opiskelijoiden välillä on englanninkielisen lyhenteen ja suomenkielisen substantiivin muodostamien yhdyssanojen tai sanaliittojen poiminnassa. Vähintään seitsemän kääntämisen opiskelijaa valitsi 21 asiantuntijoiden heti valitsemista 22:sta tämän ryhmän termistä, kun taas vähintään seitsemän merenkulun opiskelijaa valitsi niistä vain 11. Sen sijaan englanninkielisten lyhennetermien ja konjunktion *ja* sisältävien sanaliittojen valinnassa opiskelijaryhmät olivat yksimielisiä. Kummankin opiskelijaryhmän koehenkilöt valitsivat kaikki kuusi asiantuntijoiden heti valitsemaa lyhennettä ja kolme *ja*-konjunktion sisältävää sanaliittotermiä.

Suomalaiset ydintermit näyttäisivät siis olevan 1–2 sanan mittaisia substantiiveja, yhdyssubstantiiveja, lyhenteitä tai näiden yhdistelmiä, joiden frekvenssi tekstissä on vähintään kolme. Ne noudattavat siis suomalaisten termien peruskaavoja, mikä on ymmärrettävää, koska termit ovat lyhyitä. Suomalaiset asiantuntijat valitsivat lopulliseen termilistaan vain neljä sellaista vähintään kolme kertaa esiintyvää termiä, joita he eivät valinneet vielä ensimmäisellä kierroksella.

Venäläiset ydintermit ovat asiantuntijoiden poiminnan perusteella 1–2-sanaisia adjektiivin ja substantiivin sanaliittoja, substantiiveja tai lyhenteitä. Myös venäläisten ydintermien frekvenssi tekstissä on vähintään kolme. Suomalaisista asiantuntijoista poiketen venäläiset asiantuntijat eivät valinneet kaikkia suurifrekvenssisiiä termejä heti ensimmäisellä kierroksella. Ero selittynee ainakin osittain sillä, että venäläisten asiantuntijoiden poiminnassa heti valittujen termien osuus on vain kolmannes lopullisesta termimäärästä. Suomalaisen asiantuntijoiden poiminnassa heti valittujen osuus on reilusti yli puolet.

Seuraavaksi tarkastelen terminpoimintaohjelmien tuottamista termiehdokaslistoista asiantuntijoiden heti valitsemien termien saamia termiyspisteitä. Tavoitteena on selvit-

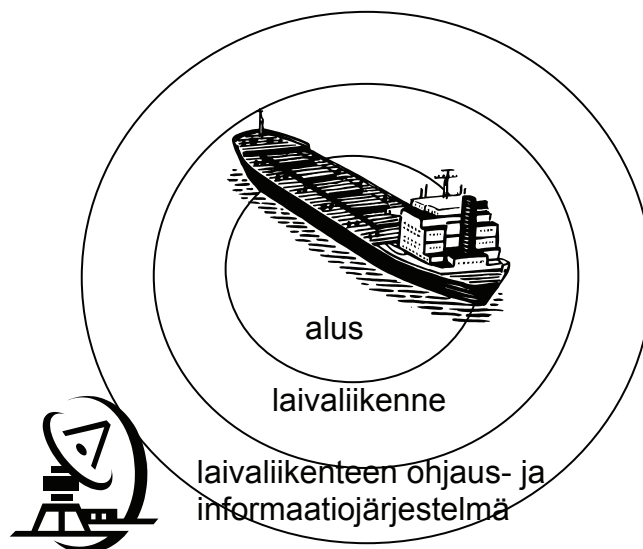
tää, ovatko asiantuntijoiden heti valitsemat termit saaneet korkeampia termiyspisteitä kuin muut termiehdokkaat, sillä erään aiemman tutkimuksen mukaan termejä on suhteessa eniten korkeita termiyspisteitä saaneiden termiehdokkaiden joukossa (Laakso 2003: 50). Terminpoiminnassa termiyspisteillä ei sinänsä ole suurta merkitystä, sillä terminologi on niistä riippumaton.

NaviTerm 2.0 -terminpoimintaohjelman tuottamassa termiehdokaslistassa on 92 niistä 118 termistä, jotka suomalaiset asiantuntijat valitsivat heti yksimielisesti termilistaansa. Näiden termien saanti on varsin korkea, 78 prosenttia. Mutta ovatko asiantuntijoiden heti valitsemat termit terminpoimintaohjelman algoritmin mukaisia vahvoja termiehdokkaita? Termit ovat saaneet keskimäärin 0,1788 termiyspistettä, kun maksimi on yksi. Yli 0,5 termiyspistettä on saanut yhdeksän termiä ja yli 0,2 termiyspistettä noin kolmannes termeistä. Noin puolet asiantuntijoiden ensimmäisellä kierroksella valitsemista termeistä on kuitenkin saanut alle 0,1 termiyspistettä. MultiTerm Extract -terminpoimintaohjelma poimi 34 termiä suomalaisten asiantuntijoiden heti ensimmäisellä kierroksella valitsemista termeistä. Ohjelman saanti näistä termeistä on 28,8 prosenttia eli vain 3,4 prosenttia parempi kuin saanti lopullisesta termilistasta. Termiyspisteet vaihtelevat 60–99 pisteeseen ja keskimäärin tämän ryhmän termit ovat saaneet 77 termiyspistettä.

Tarkastelun tulos on, että termilistassa on muutama termi, joiden termiyydestä kaikki käsinpoimijat olivat yksimielisiä ja jotka ovat myös saaneet korkeat termiyspisteet puoliautomaattisessa poiminnassa, mutta lyhenteiden lisäksi on vaikea erottaa mitään kvantitatiivisilta tai syntaktisilta ominaisuuksiltaan selkeästi muista erottuvaa ydintermien joukkoa. Hyvin yksimielisesti valitut termit ovat muutamaa poikkeusta lukuun ottamatta 1–2 sanan mittaisia. Niiden joukossa on suurifrekvenssisia termejä, mutta myös termejä, jotka esiintyvät tekstissä vain kerran tai eivät yhtään kertaa siinä muodossa, jossa ne ovat termilistassa. Frekvenssi ei siis voi olla ainoa kriteeri, jolla termiehdokkaat valitaan. Myöskään termikaavaltaan ydinjoukkoon kuuluvat termit eivät eroa muista termeistä mitenkään erityisesti. Määrälliset ja kielen rakenteeseen liittyvät perusteet ovat siis riittämättömät termiyyden arvioinnissa. Erikoisalan asiantuntijat ja koehenkilöt tekevätkin arvioinnin ilmeisesti leksikaalisten yksiköiden semanttisten ominaisuuksien ja kontekstin perusteella. Tässä konteksti on ymmärrettävä laajasti, niin että siihen kuuluu ympäröivä teksti, erikoisala ja jopa ympäröivä maailma. Poimintatilanteessa poimija arvioi muun muassa leksikaalisten yksiköiden kuulumista erikoisalan diskurssiin, niiden relevanttisuutta erikoisalatiedon kannalta ja taustalla olevan reaali maailman tarkoitteen kuulumista erikoisalan kontekstiin (vrt. Karihalme 1996: 81). Monimutkaisten kognitiivisten prosessien kautta poimijat päätyvät valitsemaan ensinnäkin lyhenteet. Koska lyhenteet ovat ensisijaisia, valitaan myös yhdyssanat tai sanaliitot, joiden osia lyhenteet ovat. Näiden lisäksi poimitaan suurifrekvenssisia substantiiveja, jotka nimeävät erikoisalan keskeisiä käsitteitä. Osa keskeisten käsitteiden nimityksistä saattaa kuulua myös yleiskieleen (esimerkiksi *alus*, *laiva*, *liikenne*, *satama*, *onnettomuus*, *turvallisuus*). Osa nimityksistä esiintyy tekstissä vain kerran tai kaksi kertaa, mutta maailmantietoa ja erikoisalatietoa hyödyntävä poimija tietää niiden kuuluvan erikoisalan käsitteistöön (esimerkiksi *luotsi*). Keskeisiä käsitteitä nimeävät termit ovat vanhoja ja vakiintuneita, eikä niissä esiinny variaatiota.

Ydinjoukon lisäksi koehenkilöt valitsivat termejä, jotka saattavat olla erilaisia kielelliseltä muodoltaan, mutta jotka tästä huolimatta saattavat nimetä samoja käsitteitä tai lähikäsitteitä. Nämä termit nimeävät erikoisalan käsitteitä, mutta kaikki niiden eri nimitykset eivät päätyneet koehenkilöiden termiehdokaslistaan. Kolmanteen termiehdokkaiden joukkoon kuuluvat termit, joiden termiyydestä asiantuntijat joutuivat neuvottelemaan ja jotka vain muutama saman ryhmän opiskelija valitsi.

Erikoisalan keskeiset käsitteet, joiden nimityksiä kutsun ydintermeiksi, kiistattomat erikoisalan käsitteet, joita nimeävissä termeissä synonymia ja variaatiota on tavallista, sekä erikoisalan ja lähialojen rajamaastoon sijoittuvat käsitteet muodostavat erikoisalan käsitekehysten. Asiantuntijoiden neuvottelua vaatineet termit nimeävät käsitteitä, jotka sijoittuvat käsitekehysten reunoille, kun taas vanhat ja vakiintuneet termit nimeävät erikoisalan selkeitä ydinkäsitteitä, jotka ovat käsitekehysten keskellä. (Vrt. Ivina 2003; Strehlow 1997.)



KUVIO 6. Merenkulun turvallisuuden käsitekehys ja ydintermit

### 6.8.7 Puoliautomaattisen terminpoiminnan etuja ja haittoja

Puoliautomaattiset terminpoimintamenetelmät ovat houkuttelevia, koska niiden avulla voidaan lyhentää terminpoimintaan kuluva aikaa. Sanastotyötä tekevä terminologi hyötyy tekstin manuaalisesta tarkastelusta alussa, jos erikoisala on outo, mutta koska sanastotyö vaatii suuren tekstimassan läpikäymistä, jossakin vaiheessa tekstit alkavat toistaa itseään, jolloin terminologi todennäköisesti haluaisi automatisoida terminpoimintaa.

Terminpoimintakokeessa käytetyistä ohjelmista NaviTerm 2.0 ja Masterin ovat yhdistelmämenetelmiä, jossa lingvististen terminpoimintamenetelmien puutteita on korjattu yhdistämällä lingvistiseen analyysiin tilastollinen analyysi (ks. jaksot 4.1.2 ja 6.7.1.1). Ohjelmien tuottamien termiehdokaslistojen saanti onkin näin saatu nousemaan manuaalisen poiminnan tasolle, mutta alhainen tarkkuus on edelleen ohjelmien heikko kohta. Alhainen tarkkuus vaikuttaa ohjelmien käytettävyyteen, koska termiehdokaslistan puhdistaminen valetermeistä vaatii runsaasti aikaa. Terminpoimintakokeen perusteella yksisanaisten yksifrekvenssisten termiehdokkaiden poistaminen ratkaisisi osittain roskaongelman, mutta samalla menetettäisiin kelvollisia termejä (vrt. Pasanen 2005b). NaviTerm 2.0 ja Masterin-ohjelmat poimivat kuitenkin hyvin yksisanaiset suurifrekvenssiset termit, joten ohjelmat ovat parhaimmillaan ydintermien poiminnassa, jolloin termiehdokkaiden ja valetermien määrän voi rajoittaa kohtuulliseksi.

MultiTerm Extract -ohjelmassa termintunnistus perustuu pääasiallisesti tilastolliseen analyysiin, mutta termiehdokaslistassa on merkkejä myös jonkinlaisesta lingvistisestä analyysistä. Tilastolliseen menetelmään perustuvat puoliautomaattiset terminpoimintaohjelmat tuottavat toistaiseksi runsaasti valetermejä, joilla on lähtötekstissä korkea frekvenssi, mutta teksteissä harvoin esiintyvät termit ja termivariantit jäävät poimimatta.



MultiTerm Extract -ohjelman ongelmia ovatkin sekä ali- että ylipoiminta, eikä niitä voida ratkaista ohjelman asetuksia muuttamalla. Suurin osa roskasta on yksisanaisia termiehdokkaita, joita ohjelma näyttää poimivan aivan liikaa. Roskaa pahempi ongelma on kuitenkin alipoiminta, jota ei pysty korjaamaan edes muuttamalla ohjelman perusasetuksia niin, että roskan määrä poimintatuloksessa lisääntyy.

MultiTerm Extract -ohjelman termiehdokaslistassa roskan määrää kasvattavat eripuiset sanayhdistelmät, joita ohjelma tarjoaa termeiksi (esimerkiksi *ympäri näköpiirin*, *ympäri näköpiirin näkyvä*, *ympäri näköpiirin näkyvä valo* ja *ympäri näköpiirin näkyvä valkoinen valo* tai *морских администраций портов*, *морские администрации*, *администраций* ja *морской*). Kahden sanan termiehdokkaat saavat aina korkeammat pisteet kuin sen osat, vaikka usein myös osat, kuten adjektiivimääritteet, ovat listassa yksisanaisina termiehdokkaina. Usein yhdessä esiintyvät sanat siis ovat ohjelman algoritmin mukaan hyvin todennäköisiä termejä. Listaan päättyy näin paljon myös sellaisia sanayhdistelmiä, jotka eivät ole termejä (esim. *c kappaleessa määrätty*, *erityisten määräysten*, *metrin alus*, *näiden sääntöjen* ja *otettava huomioon* tai *точки зрения* ja *может служить*).

Tutkimukseni tulosten perusteella MultiTerm Extract -ohjelma on käyttökelpoinen apu terminpoiminnassa, jos riittää, että tekstistä poimitaan vain erikoisalan keskeisiä käsitteitä nimeävät ydintermit. Sanastotyössä MultiTerm Extract -ohjelman tasoisesta automaattisesta terminhausta on ehkä eniten hyötyä, kun terminologin on ratkaistava, sisältääkö luokituksen tai otsikon perusteella aineistoon valittu teksti valitun alan termejä. Automaattisen terminhaun hakutuloksessa on kuitenkin niin paljon puutteita, että ainoaksi terminhakumenetelmäksi se ei sovellu.

Sulkulistaa hyödyntävissä terminpoimintaohjelmissa, kuten Tradoksen MultiTerm Extract, olisi mahdollista vähentää huomattavasti roskan määrää täydentämällä sulkulistaa. Sulkulistojen käytön ongelma on, että ne laaditaan yleensä tiettyä aineistoa ja tiettyä tarkoitusta varten eli ne ovat kertakäyttöisiä. Kaikkiin aineistoihin sopivaa sulkulistaa on vaikea laatia. Suodattimena se on joko liian tiheä tai liian harva. MultiTerm Extract -ohjelman sulkulista onkin liian yleinen. Suomen sulkulistalla on kaikkiaan 398 sanaa, mutta jos sanat palautetaan perusmuotoon, jää jäljelle noin 200 eri sanaa. Listalla ovat pronominit, konjunktiot, yleisimmät adverbis ja muun muassa *olla*-verbi. Sulkulistaa voisi täydentää adverbeilla (*kuitenkin*, *kuten*) ja verbeillä, jotka eivät voi olla termejä (*tulla*, *voida*). Myös tavalliset lyhenteet, kuten *esim.* ja *jne.*, voisivat olla sulkulistalla. Venäjän kielen sulkulista on pidempi, siinä on 587 sanaa, mutta sekin lyhenee alle puoleen, kun sanat palautetaan perusmuotoon. Erilaisten taivutusmuotojen runsaus on haaste sulkulistan laatijalle yhtä hyvin venäjän kuin suomen kielessä.

Ohjelmointitekniisten haasteiden ohella terminpoimintaohjelmien käyttöön liittyy ongelmia. Ohjelmat ovat usein tietynkielisten aineistojen käsittelyä ja tiettyä tutkimusprojektia varten kehitettyjä. Myös käyttöoikeuskysymykset rajoittavat niiden laajempaa käyttöä. Tällä hetkellä sopivan ohjelman löytäminen suurempi ongelma on kuitenkin sopivan aineiston löytäminen. Vaikka tekstejä on sähköisessä muodossa, puoliautomaattiseen terminpoimintaan sopivan korpuksen kokoaminen vaatii aikaa ja käsityötä, olipa korpuksen käyttötarkoituksena sanastotyö tai terminologinen tutkimus.

Terminpoimintakokeeni perusteella puoliautomaattisia terminpoimintaohjelmia kannattaa käyttää ydintermien poimintaan. Ydintermien poiminnan jälkeen terminologisen tiedon poimintaa jatketaan muilla menetelmillä, joista seuraavaksi esittelen koettimien käyttöön perustuvan termien ja terminologisen tiedon poimintamenetelmän.



## 7 KÄSITESUHTEIDEN POIMINTA TERMINOLOGISTEN KOETTIMIEN AVULLA SUOMALAISISTA JA VENÄLÄISISTÄ MERENKULUN TURVALLISUUTTA KÄSITTELEVISTÄ TEKSTEISTÄ

Terminologisesti kiinnostavan tiedon poiminnan automatisoinnissa on viime aikoihin saakka keskitytty terminpoiminnan automatisointiin. Korpusmenetelmien kehittyessä huomattiin, että termin kontekstissa voi olla muutakin terminologisesti kiinnostavaa tietoa, kuten käsitteen määritelmä, selite tai käsitepiirteen kuvaus (Bowker & Pearson 2002: 38).

### 7.1 Käsitetieto teksteissä

Käsitetiedolla tarkoitan tietoa käsitteiden välisistä suhteista ja käsitepiirteistä. Näitä tietoja tarvitaan käsitteiden määrittelyssä. Käsitetieto ei kuitenkaan ole sama asia kuin määritelmätieto, vaan tätä laajempi käsite. Terminologisessa ihannemääritelmässä käsitteestä annetaan minimimäärä tietoa, jonka avulla se erotetaan muista käsitteistä. Käsitetieto sen sijaan voi sisältää sellaistaakin tietoa käsitteestä, joka ei ole tarpeen käsitteen määrittelemiseksi. *Käsitetiedon* käsitteen ymmärtämiseksi selostan ensin tutkijoiden näkemyksiä erilaisista määritelmistä.

Perinteinen terminologia hyväksyy terminologiseksi määritelmiksi vain sisältömääritelmän ja joukkomääritelmän (vrt. Sanastotyön käsikirja 1988: 41). Yläkäsitteen nimityksen ja erottavan piirteen muodostamaa sisältömääritelmää pidetään perinteisen terminologian mukaan määritelmän ihannetapauksena. Tällaisissa määritelmissä, joita Temmerman (1996: 829) kutsuu ydinmääritelmiksi (core definition) ja Pearson (1998: 102) formaalisiksi määritelmiksi (formal definition), on vain minimimäärä tietoa käsitteen erottamiseksi muista käsitteistä. Tähän tarvitaan käsitesuhde, jonka avulla käsite voidaan sijoittaa omalle paikalleen käsitejärjestelmässä, ja yksi olennainen käsitepiirre, joka erottaa määriteltävän käsitteen sen yläkäsitteestä ja vieruskäsitteistä. Tekstissä esiintyessään sisältömääritelmä on yleensä muodossa  $X = Y + \text{erottava piirre}$  tai muodossa  $Y + \text{erottava piirre} = X$ . Lausekkeessa  $X$  on määriteltävän käsitteen nimitys ja  $Y$  yläkäsitteen nimitys. Erottava piirre ilmaistaan relatiivilauseella tai muulla määritteellä. (Vrt. Pearson 1998: 102.) Esimerkiksi käsite *konealus* voitaisiin määritellä seuraavasti: alus, jota kuljettaa koneisto.

Toinen terminologisessa kirjallisuudessa mainittu terminologinen määritelmä on joukkomääritelmä. Siinä luetellaan kaikki määriteltävän käsitteen lähimmät alakäsitteet tai käsitteen alaan kuuluvat tarkoitteet. Esimerkiksi käsitteen *maaliikenne* joukkomääritelmä voisi olla: maantieliikenne ja raideliikenne. Tässä määritelmässä ovat yläkäsitteen *maaliikenne* lähimmät geneeriset alakäsitteet. Alakäsitteitä koskee vaatimus, että niiden tulee olla toisensa pois sulkevia. Joukkomääritelmän etu on, että se ei ole niin analyttinen kuin sisältömääritelmä ja joukkomääritelmää voi käyttää, vaikka käsitesisältöä ei olisi tutkittu tarkoin, koska joukkomääritelmässä ei anneta käsitepiirteitä. Määritelmätyypin puute on, että alakäsitteiden luettelo voi olla pitkä.

Terminologiset määritelmät ovat ihanteellisia käsitetiedon poiminnan kannalta, sillä niissä käsitteet määritellään toisten käsitteiden avulla siten, että käsitesuhteet ilmenevät määritelmistä (Šelov 2001: 8). Ne eivät kuitenkaan ole kovin produktiivisia, sillä erikoisalatekstien kirjoittajat määrittelevät käsitteitä terminologisten määritelmien avulla vain harvoin. Jos tekstejä aineistonaan käyttävä terminologi tyytyisi hakemaan esimerkiksi vain sisältömääritelmiä, jäisi suuri osa tekstien sisältämästä käsitetiedosta keräämättä. Sisältö- ja joukkomääritelmien lisäksi teksteissä on kontekstimääritelmiä, jotka eivät täytä sisältö- tai joukkomääritelmän muotovaatimuksia, mutta esimerkiksi Šelov

(2001: 3) pitää myös kontekstimääritelmiä terminologisina määritelminä. Šelovin (1998: 43) mukaan kontekstimääritelmät ovat yleisiä erityisesti humanistisilla aloilla. Antonova (1984: 106) käyttää vastaavista määritelmistä nimitystä *epäsuora määritelmä* (непрямое, неспециальное определение). Hänen mukaansa epäsuora määritelmä voi olla lauseen osa, eikä lauseen varsinainen tarkoitus ole määritelmän kertominen toisin kuin suorassa määritelmässä (прямое, специальное определение), jossa kirjoittajan tarkoitus on nimenomaan käsitteen määrittely.

Sisältö-, joukko- ja kontekstimääritelmien lisäksi teksteissä voi olla määritelmiä, joissa ei anneta erottavaa piirrettä mutta kylläkin yläkäsitteen nimitys tai joissa annetaan erottava piirre mutta ei yläkäsitettä. Tällaiset käsitetiedon muruset muodostavat vajaita määritelmiä, joissa on tietoa käsitesuhteista tai käsitepiirre, mutta ei kumpaakin. Pearson (1998: 135) käyttää näistä määritelmistä nimityksiä *puolimuodollinen* (semi-formal) ja *ei-muodollinen* (non-formal) *määritelmä*. Puolimuodollisessa määritelmässä annetaan esimerkiksi erottava piirre, mutta ei yläkäsitettä, kuten Pearsonin (1998: 99) esimerkkilauseessa ”An anemometer registers the speed of the wind on a dial or gage”. Ei-muodollisesta määritelmästä Pearson (1998: 80) antaa esimerkin ”an arachnid is a spider”. Mielestäni tässä esimerkissä on annettu käsitteen luokka, mutta ei erottavaa piirrettä. Tutkimuksessani haen suomalaisia ja venäläisiä koettimia kaikille edellä kuvatuille määritelmätyypeille.

Käsitteen nimitys saatetaan tekstissä korvata synonyymillä, lähikäsitteen nimityksellä, selittävällä ilmauksella tai pronominilla. Pearson (1998: 168) pitää näitäkin lausumia tärkeinä käsitetiedon lähteinä, sillä niistä on hyötyä määritelmien muotoilussa (ks. myös Haarala 1981: 44). Jo yksistään synonyymien tai tietyssä kontekstissa käytettävän oikean termin ilmoittaminen on arvokas tieto (Pearson 1998: 119). Šelov (2001: 6) sen sijaan varoittaa erottelemattomista määritelmistä (неспецифическое определение), jotka eivät anna käsitteelle sen paremmin erottavaa piirrettä kuin yläkäsitettäkään ja joista ei näin ollen ole hyötyä käsitteen sijoittamisessa käsitejärjestelmään. Esimerkiksi Šelov (2001: 6) antaa käsitteen *heating time* määritelmän *length of time required for heating*, joka ei anna tietoa käsitteen paikasta käsitejärjestelmässä.

Määritelmien lisäksi teksteissä on paljon muutakin käsitetietoa, joka voi olla hajallaan, jopa piilossa. Teksteissä voi olla ensyklopedista tai ideologisesti värittynyttä tietoa, joka ei ole välttämätöntä terminologisen määritelmän muodostamiseksi, mutta auttaa lukijaa saamaan kokonaiskuvaa erikoisalan käsitteistöstä. Käsitetieto onkin mielestäni huomattavasti laajempi käsite kuin vain määritelmän muodostamiseksi tarpeelliset käsitesuhteet ja käsitepiirteet. Tutkimuksessani käsitetieto on ymmärretty laajasti, niin että sen alaan kuuluu kaikenlainen terminologisesti kiinnostava tieto, myös termejä koskeva. Feliu ja Cabré (2002) käyttävät tällaisen terminologisesti kiinnostavan tiedon osasta nimitystä erikoisalatiedon yksikkö (specialized knowledge unit).

Olen edellä kuvannut, millaista käsitetietoa erikoisalateksteissä mahdollisesti on. Seuraavaksi tarkastelen lähemmin käsitesuhteita ja käsitepiirteitä, jotka ensisijaisesti kiinnostavat terminologia, kääntäjää, tulkkia, teknistä kirjoittajaa tai erikoisalan asiantuntijaa silloin, kun tavoitteena on hahmottaa erikoisalan käsitejärjestelmää.

### 7.1.1 Käsitesuhteet

Käsitteet eivät ole irrallisia, vaan aina suhteessa toisiin käsitteisiin. Standardin ISO 704 (2000: 5–12) mukaan kahden käsitteen välillä voi olla hierarkkinen geneerinen suhde<sup>1</sup> (generic relation), hierarkkinen koostumussuhde (partitive relation) tai funktiosuhde (as-

<sup>1</sup> Geneeristä käsitesuhdetta nimitetään suomessa myös sisältösuhteeksi.

sociative relation). Kahden käsitteen välillä vallitsee geneerinen suhde, jos alakäsitteen käsitesisältö sisältää yläkäsitteen käsitesisällön ja sen lisäksi vähintään yhden erottavan piirteen. Kaksi käsitettä on koostumussuhteessa toisiinsa, jos yläkäsite edustaa kokonaisuutta, jonka osia alakäsitteet ovat. Kolmatta standardissa mainittua käsitesuhdetta, funktiosuhdetta, on pidetty eräänlaisena kaatoluokkana. Jos kahden käsitteen välillä ei ole geneeristä suhdetta eikä koostumussuhdetta, niiden välillä on jonkinlainen funktiosuhde, jota ei välttämättä ole selvitetty tarkemmin. Funktiosuhteen määritelmä on yhä vielä epämääräinen. ISO 704 -standardin (2000(E): 12) mukaan käsitteiden välillä on funktiosuhde, jos kokemuksen perusteella niiden välillä on temaattinen yhteys. Geneerisistä käsitesuhteista ja koostumussuhteista funktiosuhteet eroavat siten, että funktiosuhteessa käsitteet eivät ole hierarkkisessa suhteessa toisiinsa. Kumpikaan käsitteistä ei siis ole toisen yläkäsite.

Toinen yleinen näkemys on, että kahden käsitteen välillä voi olla looginen tai ontologinen käsitesuhde. Looginen käsitesuhde on suhde, joka voidaan havaita vertailemalla käsitteiden sisältöä eli intensiota ja alaa eli ekstensiota (Nuopponen 1994: 82, 84). Loogiset käsitesuhteet ovatkin varsinaisia käsitteiden välisiä suhteita ja perustuvat niiden laadulliselle vertailulle käsitetasolla (Pilke 2000: 60). Looginen käsitesuhde on esimerkiksi sisältyminen (inklusion) yläkäsitteeseen eli ISO 704 -standardin geneerinen suhde (ks. esim. Nuopponen 1994: 114). Ontologinen käsitesuhde taas on suhde, joka on riippuvainen reaalimaailman tarkoitteiden välisistä suhteista (Nuopponen 1994: 69; Pilke 2000: 59). Ontologiset suhteet viittaavat tarkoitteiden todellisiin suhteisiin reaalimaailmassa ja perustuvat ilmiöiden (havaittaviin) eroihin. Niissä tavallaan simuloidaan reaalimaailmaa (Pilke 2000: 60). Ontologisia käsitesuhteita on lukuisia. Nuopponen (1994: 86–87, 98–99) jakaa ontologiset käsitesuhteet kontaktisuhteisiin (kontaktrelationer) ja vaikutussuhteisiin (influeringsrelationer). Kontaktisuhteessa käsitteiden välillä on paikkaan tai aikaan sidottu yhteys, kuten esimerkiksi pöydän ja pöydänjalan tai kylvämisen ja koulumisen välillä. Vaikutussuhteessa taas käsitteet vaikuttavat yksi- tai kaksisuuntaisesti toistensa olemassaoloon. Vaikutussuhteessa yhteys on siis ennen kaikkea kausaalinen. (Nuopponen 1994: 87–111.)

Nuopposen luokittelussa kontaktisuhteisiin kuuluva koostumussuhde (partitiv begreppsrelation) on suppeampi kuin ISO 704 -standardin mukainen koostumussuhde, sillä Nuopposen luokittelussa koostumussuhteisten käsitteiden välillä on paikkaan sidottu yhteys, kun taas ajallisesti toisiinsa liittyvien käsitteiden välillä on temporaalinen käsitesuhde. Standardin mukaisessa luokittelussa jotkut ajallisesti toisiinsa liittyvien käsitteiden välisistä suhteista ovat koostumussuhteita. Tällaisia ovat esimerkiksi prosessit, joiden voidaan ajatella koostuvan vaiheista, jolloin prosessi on koostumussuhteen yläkäsite ja vaiheet sen alakäsitteitä. Koostumussuhteisen ja temporaalisen käsitejärjestelmän ero on se, että koostumussuhteisessa käsitejärjestelmässä on periaatteessa samantekevää, missä järjestyksessä saman tason alakäsitteet ovat käsitejärjestelmää kuvaavassa kampadiagrammissa, kun taas temporaalisessa käsitejärjestelmässä prosessin vaiheet seuraavat toisiaan tietyssä järjestyksessä (vrt. Nuopponen 1994: 176).

ISO 704 -standardin noudattama perinteinen jako käsitesuhteisiin ei ole likikään niin tarkka kuin Nuopposen jaottelu. Tästä syystä suurin osa Nuopposen luettelemista ontologisista käsitesuhteista kuuluu perinteisessä jaottelussa aiemmin mainittuun kaatoluokkaan eli funktiosuhteisiin. Nuopposen väitöskirja (1994) onkin ehkä perusteellisin käsitesuhteita koskeva tutkimus ja antaa työkaluja myös muiden kuin geneeristen käsitesuhteiden tarkasteluun.

Feliu ja Cabré (2002) ovat esitelleet käsitesuhteiden luokittelun, joka on tarkoitettu käsitesuhteiden poimintaan teksteistä. Luokittelu vastaa osittain ISO-standardin ja Nuopposen käsitesuhteiden luokittelua, mutta myös eroaa niistä joiltakin osin. Feliun ja

Cabrén mukaan käsitteiden välillä voi olla geneerinen suhde tai koostumussuhde, joista he käyttävät nimityksiä *inclusion* ja *meronymy*. Lisäksi heidän luokittelussaan on viisi muuta käsitesuhdetta: samanlaisuus (similarity), peräkkäisyys (sequenciality), kausaalisuus (causality), instrumentaalisuus (instrument) ja assosiaatio (association). Samanlaisuuden käsitesuhteessa käsitteet ovat täysin tai osittain samat tai täysin tai osittain erilaiset. Käsitteiden samanlaisuutta ei yleensä ole pohdittu käsitesuhteiden yhteydessä, vaan ainoastaan termien poiminnan yhteydessä, sillä synonymia on tärkeä termien välinen suhde.

Peräkkäisyyttä ilmaiseva käsitesuhde vastaa Nuopposen kontaktisuhdetta. Nuopposen tavoin Feliu ja Cabré erottavat aikaan ja paikkaan liittyvän käsitteiden välisen yhteyden. Kausaalinen ja instrumentaalinen suhde ovat Feliun ja Cabrén luokittelussa omana käsitesuhdetyyppinä. Lisäksi tutkijat erottavat erilliseksi käsitesuhteeksi assosiaation, joka käsitteiden välillä on silloin, kun käsitteet liittyvät toisiinsa jollakin tarkemmin määrittelemättömällä tavalla. Feliun ja Cabrén assosiaatiolle esittämä määritelmä on hyvin lähellä ISO 704 -standardin mukaisen funktiosuhteen määritelmää. Yhteinen piirre määritelmissä on niiden epämääräisyys.

Tämän tutkimuksen tarpeisiin sanastotyössä laajalti käytetty ISO 704 -standardin mukainen käsitesuhteiden luokittelu geneerisiin, koostumus- tai funktiosuhteisiin on liian karkea, niin kuin varmasti lähes kaikkiin terminologisiin tutkimustarkoituksiin. Nuopposen luokittelu puolestaan on ehkä liiankin tarkka, joten sovellan tässä tutkimuksessa näiden luokittelujen välimuotoa. Geneeristen käsitesuhteiden luokittelussa noudatan ISO 704 -standardin geneerisen suhteen määritelmää, jonka mukaan alakäsitteellä on yläkäsitteen sisältö ja lisäksi vähintään yksi erottava piirre. Standardin mukainen geneerinen käsitesuhde vastaa Nuopposen loogista käsitesuhdetta. Koostumussuhteita tarkastelen standardin ISO 704 ja lähinnä Sanastokeskuksen luoman suomalaisen sanastotyöperinteen mukaisesti. Käsitän koostumussuhteen laajemmin kuin Nuopponen ja sisällytän koostumussuhteisiin käsitesuhteet, joissa yläkäsite edustaa kokonaisuutta, jonka osia alakäsitteet ovat, silloinkin kun käsitteiden välillä on paikkaan tai aikaan sidottu yhteys. Luen koostumussuhteisiin esimerkiksi prosessit, jotka koostuvat ajallisesti toisiaan seuraavista vaiheista.

Ylä- ja alakäsitteen välisen käsitesuhteen lisäksi geneerisiä käsitesuhteita tai koostumussuhteita sisältävässä käsitejärjestelmässä on toinenkin tutkimukseni kannalta tärkeä käsitesuhde. Tämä on vieruskäsitesuhde, jossa kahdella käsitteellä on sama lähin yläkäsite. Lisäksi vieruskäsitesuhteen ehtona on, että käsitteet sijoittuvat samaan käsitejärjestelmän ulottuvuuteen. (Terminologian sanasto, s.v. *vieruskäsite*.)

Funktiosuhteiden luokittelussa olen käyttänyt lähtökohtana Nuopposen jaottelua kontakti- ja vaikutussuhteisiin (kontaktrelation; influeringsrelation). Nuopposen luokittelussa kontaktisuhdetta käsitteiden välillä on paikkaan tai aikaan sidottu yhteys. Omasta luokittelustani olen jättänyt kontaktisuhdetta luokan kokonaan pois luokittelemalla koostumussuhteisiin käsitesuhteet, joissa käsitteet ovat kokonaisuuden osia, ja vaikutussuhteisiin käsitesuhteet, joissa käsitteiden välillä on kausaalinen yhteys. Vaikutussuhteita ovat toiminnan ja sen tarkoituksen tai tuloksen välinen sekä tapahtuman ja sen syyn tai seurauksen väliset käsitesuhteet. Myös vaikuttava tekijä on vaikutussuhteessa toiseen käsitteeseen. Muita funktiosuhteita ovat esimerkiksi toiminnan ja toimijan tai toiminnan kohteen väliset käsitesuhteet. Samojen käsitteiden välillä voi vallita monenlaisia suhteita, mutta ratkaisevaa on, mikä suhde on tärkein siinä kontekstissa, jossa käsitteistä puhutaan. Esimerkiksi syyn ja seurauksen välillä ratkaiseva on kausaalisuus eikä aika, vaikka syy ja seuraus liittyvätkin toisiinsa ajallisesti.

### 7.1.2 Käsitepiirteet

Terminologisessa kirjallisuudessa käsitepiirteet luokitellaan yleensä sisäisiin ja ulkoisiin sekä olennaisiin (välttämättömiin, määrittäviin) ja täydentäviin (epäolennaisiin, lisä-) piirteisiin (ks. esim. ISO 704:2000(E): 4–5; Pilke 2000: 56–57; Sanastotyön käsikirja 1988: 27). Sisäiset piirteet liittyvät itse tarkoitteeseen kun taas ulkoiset piirteet liittyvät tarkoitteiden keskinäiseen vertailuun (Pilke 2000: 56–57). Sisäisiä piirteitä ovat esimerkiksi muoto tai koko, ulkoisia ovat muun muassa käyttötarkoitus tai valmistaja. Esimerkiksi matkustaja-aluksen käyttötarkoitusta ilmaiseva ulkoinen käsitepiirre on, että se on tarkoitettu matkustajien kuljettamiseen. Tämä on myös olennainen piirre, jos luokittelu-  
perusteena käytetään kuljettettavaa lastia.

Terminologisia menetelmiä ja piirreluokitteluja on kehitetty lähinnä konkreettisten oliokäsitteiden tarkasteluun, eikä niitä voi suoraan soveltaa muunlaisiin käsitteisiin (Pilke 2000: 320; ks. jakso 2.1). Huomattava osa erikoisalojen käsitteistä on kuitenkin dynaamisia toiminta- tai tapahtumakäsitteitä. Pilkkeen mukaan (Pilke 2000: 109) tekniikan, lääketieteen ja oikeustieteen sanastoissa dynaamisten käsitteiden osuus on 13–19 prosenttia. Pilkkeen dynaamisten käsitteiden piirteitä käsittelevä väitöskirja täydentääkin terminologista tutkimusta tärkeältä terminologian osa-alueelta. Karihalme (1996: 91) puolestaan huomasi muotoilun teorian sanastoa tutkiessaan, että terminologisesta kirjallisuudesta saa niukasti tukea ominaisuuksikäsitteiden tarkasteluun. Ominaisuuksien tarkastelu pikemmin käsitepiirteinä kuin käsitteinä. Ominaisuuksikäsitteiden tutkimuksessa Karihalme tukeutuikin filosofiaan ja kognitiotieteisiin, koska terminologian teoria ei tarjonnut riittävästi teoreettista tukea.

Käsitteitä on myös tarkasteltu käsitejärjestelmästä irrallisina sen sijaan, että piirteiden tarkasteluun olisi alakohtaiset ja käsiteluokkakohdaiset mallit (Pilke 2002: 12). Koska alakohtaisia ja käsiteluokkakohdaisia malleja ei ole, tarkoitteiden ominaisuuksista havaitaan ja valitaan ne, jotka sopivat tiettyyn tarkoitukseen, esimerkiksi sanastoon tai tutkimukseen. Alakohtaisten ja käsiteluokkakohdaisien mallien kehittämiseksi on tehty joitakin tutkimuksia (ks. esim. Ivina 2003; Komarova 1991). Komarova (1991) on tutkinut sisältöominaisuuksien erottavia piirteitä maatalousalan GOST-standardeissa ja päättänyt ala- ja käsiteluokkakohdaiseen malliin, jonka avulla maatalousalan käsitteitä voidaan määrittellä. Komarova toteaa, että erottavat piirteet riippuvat käsiteluokasta eli käsitteillä on käsiteluokan mukaiset, tyypilliset erottavat piirteensä. Kolmen erikoisalan dynaamisia toiminta- ja tapahtumakäsitteitä tutkinut Pilke (2000) puolestaan päätyi tulokseen, jonka mukaan dynaamisten käsitteiden kuvaamiseen ja tutkimukseen voidaan soveltaa yleispäteviä, alasta riippumattomia periaatteita. Hänen tutkimuksensa tuotti käsiteluokkakohdaisen mallin dynaamisten toiminta- ja tapahtumakäsitteiden kuvaamiseen. Käsitepiirteiden tarkastelu käsiteluokittain olisi kiinnostavaa tutkimuskohde käsitetiedon poiminnan näkökulmasta. Tässä tutkimuksessa en aiheen laajuuden vuoksi kuitenkaan tarkastele käsitepiirteiden poimintaa lähemmin vaan keskityn käsitesuhteisiin liittyvän käsitetiedon poimintaan.

Käsitetiedon paikallistamiseksi teksteistä on kehitetty erilaisia menetelmiä, joista yksi on terminologisten koettimien käyttö paikallistamisen apuna. Terminologisten koettimien käytön teoreettista taustaa olen selostanut jaksossa 4.2. Tähänastinen tutkimus on koskenut lähinnä englannin kieltä, mutta koska koettimien toimintaperiaate on kielestä riippumaton, on perusteltua uskoa, että koettimia voi käyttää myös suomen ja venäjän kielessä. Tässä luvussa raportoin sitä tutkimukseni empiiristä osaa, jossa tavoitteena oli tunnistaa suomen- ja venäjänkielisiä koettimia.



## 7.2 Terminologiset koettimet termien ja käsitetiedon poiminnassa

Englanninkielisillä aineistoilla tehty koetintutkimus alkoi kiinnostaa tutkijoita laajemmin 1990-luvulla (esimerkiksi Ahmad & Fulford 1992; Kavanagh 1995), mutta venäjänkielisessä terminologisessa kirjallisuudessa kielelliset tai painotekniset keinot, joiden avulla voidaan ilmaista käsitesuhteita tai käsitepiirteitä, on mainittu jo 1980-luvun alkupuolella. 1990-luvulla tiedon poiminnan automatisointi nousi selkeästi koettimien tunnistamisen tavoitteeksi. Siihen pyrki esimerkiksi Kavanagh (1995), joka haki koettimia englanninkielisistä tietojenkäsittelyalan teksteistä. Hänen tarkoituksenaan oli suunnitella tietokoneohjelma, jolla teksteistä pystyttäisiin tunnistamaan terminologisesti kiinnostava tieto automaattisesti. Kavanagh'n kehittämän *Text Analyzer* -ohjelman avulla englanninkielisestä tekstistä pystytään automaattisesti tunnistamaan geneerisiä suhteita, koostumussuhteita, funktiosuhteita ja synonymiaa.

Tutkijat ovat yleensä käyttäneet lähtökohtana aiemmissa tutkimuksissa todettuja koettimia. Näin on menetelty esimerkiksi Grinsted (2000), joka haki yritystoimintaa käsittelevistä englanninkielisistä aikakauslehtiartikkeleista käsitesuhteita aluksi Kavanagh'n koettimien avulla, mutta laajensi hakua myöhemmin muihin aineistossa esiintyneisiin koettimiin. Grinstedin tavoitteena oli tunnistaa ne koettimet, jotka olivat käyttökelpoisimmat nimenomaan yritystoiminnan alalla. Uusissa tutkimuksissa todetut koettimet ovat osittain olleet samoja kuin aiemmissa tutkimuksissa havaitut, mutta yleensä luettelo on laajentunut entisestään. Näin siksi, että koettimet ovat kirjoittaja- ja aineistokohtaisia eikä mitään yksiselitteistä, kaikkiin teksteihin ja kaikille aloille sopivaa koettimien luetteloa ole olemassa.

Koettimien tunnistamista vaikeuttaa se, että hakua ei voi rajata esimerkiksi tietyn sanaluokan sanoihin tai edes sanoihin (ks. jakso 4.2). Aiempien tutkimusten perusteella monipuolisimpia koettimia ovat verbit, ja jotkut tutkijat ovatkin päättäneet keskittyä verbikoettiin (esim. Antonova 1984; Christensen 2000). Esimerkiksi Christensenin (2000) tavoitteena oli koota luettelo tanskan verbeistä, joita voidaan käyttää määritelmien poimintaan tanskankielisestä hydrauliikka-alan korpuksesta. Tämä verbien luettelo toimisi perustana sellaisten koettimien listalle, joka olisi yleispätevä tiedon poimintaan minkä tahansa erikoisalan sähköisestä korpuksesta. Christensenin tutkimuksen mukaan verbikoettimien avulla voidaan hakea ainakin geneerisiä suhteita, koostumussuhteita ja vieruskäsitteitä. Jotkut tutkijat taas ovat keskittyneet tietyn käsitesuhteen poimintaa helpottavien koettimien tutkimukseen (esim. Sisseck 2005). Yleensä tutkijat ovat olleet kiinnostuneita koettimista, jotka ilmaisevat synonymiaa, geneeristä suhdetta, koostumussuhdetta tai funktiosuhdetta. Eniten koettimien toimintaa on tutkittu geneerisen suhteen yhteydessä, ja esimerkiksi englanninkielisten geneeristä suhdetta ilmaisevien koettimien luettelo on melkoisen pitkä.

En ole rajannut tutkimustani etukäteen tietynlaisiin koettimiin, vaan lähtökohtana on koota aineistostani kaikki erilaiset ja erilaista terminologisesti kiinnostavaa tietoa ilmaisevat koettimet. Lähempään tarkasteluun olen valinnut produktiiviset, vähän roskaa tuottavat koetinehdokkaat.

### 7.2.1 Terminologisten koettimien käytön hyödyt

Terminologisten koettimien tärkein etu on, että niiden toimintaperiaate on kielestä riippumaton. Osa itse koettimistakin on kielestä riippumattomia ei-kielellisiä, graafisia merkkejä. Tällaisia ovat välimerkit, joita saatetaan käyttää terminologisen tiedon yhteydessä, kuten sulkeet, kaksoispiste tai pilkku. Vaikka suurin osa koettimista on kielestä riippuvia kielellisiä ilmauksia (Ahmad & Rogers 2001: 749–750), yritys tunnistaa sään-



nölliset kaavat, joiden avulla kirjoittajilla on tapana viitata termeihin ja käsitetietoon teksteissä, on tutkijoiden mielestä lupaava (ks. Bowker & Pearson 2002: 219).

Bowker (1996b: 35) pitää koettimia erittäin lupaavana menetelmänä, sillä hän väittää, että terminologisten koettimien avulla teksteistä voidaan nopeasti tunnistaa kohdat, joihin käsitetieto on kasautunut (knowledge-rich areas of text). Tietokasautumien perusteella lukija voi päättää, onko teksti sopiva lähempään tarkasteluun. Käsitetiedon ja termien käyttöesimerkkien paikallistamisesta on hyötyä erityisesti oudon alan tekstien tarkastelussa ja uusien termien poiminnassa (Pearson 1998: 210). Koettimet sopivat hyvin käsitetiedon poimintaan uusista erikoisalan teksteistä, koska koettimia voi käyttää myös silloin, kun aineistoa ei ole tagattu eli sanoja ei ole varustettu sanaluokkaa ilmaisevilla leimoilla. Koettimista on selkeästi hyötyä myös silloin, kun on ratkaistava onko termiehdokas termi vai ei (Pearson 1998: 125). Jos termiehdokas esiintyy yhdessä termiä osoittavan kielellisen ilmaisimen kanssa, termiehdokas on Pearsonin mielestä todennäköisesti termi.

Meyer (2001: 282) luettelee kolme koettimien käyttötarkoitusta: määritelmien poiminta, määritelmien muotoilu ja erikoisalatiedon kartuttaminen. Koettimet ovat hyödyksi jo määritelmien poiminnassa, sillä monia metakielen kaavoja voi vain vähän tietokonetta apuna käyttäen käyttää terminologisen määritelmän muotoiluun ja moni niiden avulla haettu määritelmä toimii sellaisenaan määritelmänä (Pearson 1998: 206). Meyerin mainitsemista käyttötarkoituksista merkittävämpiä ovat todennäköisesti kuitenkin toisena ja kolmantena listalla olevat määritelmien muotoilu ja erikoisalan tiedon kartuttaminen, varsinkin jos aineistona ovat aidot käyttötekstit. Koettimien avulla teksteistä voidaan poimia määritelmien muotoilussa tarpeellista käsitetietoa. Ehkä tärkein koettimien käytöllä saavutettava hyöty on, että teksteistä saadaan puoliautomaattisesti ainakin osa erikoisalatiedosta, joka aiemmin on pitänyt hankkia konsultoimalla alan asiantuntijoiden kanssa tai kahlaamalla läpi suuri määrä tekstimassaa (Pearson 1996: 824, 1998: 208).

Lupaavista tutkimustuloksista huolimatta koettimien käyttöön terminologisesti kiinnostavan tiedon poiminnassa liittyy ongelmia, joita ei toistaiseksi ole pystytty ratkaisemaan tyydyttävästi.

### **7.2.2 Terminologisten koettimien käyttöön liittyviä ongelmia**

Osa terminologisten koettimien käyttöön liittyvistä ongelmista johtuu yksinkertaisesti siitä, että kielenkäyttöä ei voi kahlita edes erikoisaloilla. Ihmiset ilmaisevat asioita moninaisin eri keinoin ja usein epämääräisestikin. Koska periaatteessa mikä tahansa sana, sanaliitto tai välimerkki voi olla koetin, koetinhakua ei voi rajata esimerkiksi tietyn sanaluokan sanoihin tai edes sanoihin (Piotrovskij ym. 1981: 36; Šilova 2005: 12). Näin mahdollisten koettimien luettelosta tulee helposti pitkä, mikä hankaloittaa automaattista terminologisen tiedon hakua. Pituudestaan huolimatta mikään mahdollisten koettimien lista ei kuitenkaan ole täydellinen (vrt. Bowker & Pearson 2002: 218). Aiemmissa tutkimuksissa (esim. Grinsted 2000: 50) on lisäksi todettu, että koettimet ovat kirjoittaja- ja alakohtaisia.

Kielen monet keinot ilmaista suhteita ovat ongelma erityisesti puoliautomaattisessa terminologisen tiedon tunnistuksessa, sillä ”kielelliset rakenteet ovat arvaamattomia, monimerkityksisiä ja alakohtaisia”. Suhteita ilmaistaan esimerkiksi välimerkeillä ja termit korvataan usein pronomineilla. (Soininen 1999a: 52.) Jotkut koettimet ovat niin sanottuja yleiskoettimia, jotka eivät yhdisty selkeästi vain tiettyyn käsitesuhteeseen tai käsitepiirteeseen, vaan ilmaisevat monenlaista terminologisesti kiinnostavaa tietoa (vrt. Soininen 1999: 56). Tällaisia ovat aiempien tutkimusten perusteella varsinkin välimerk-

kikoettimet, mutta myös verbikoettimet voivat liittyä moneen suhteeseen. Esimerkiksi englannin kielen koetin *describe\** (*as*) yhdistyy sekä geneeriseen että funktiosuhteeseen (Grinsted 2000: 47). Suhteen selvittäminen vaatii manuaalista työtä, vaikka sen vähentämiseen koettimien käytöllä nimenomaan pyritään.

Muita koettiin liittyviä ongelmia on esimerkiksi se, että käsitteen nimityksen ja käsitetiedon välissä voi olla ”ylimääräistä” tekstiä eli tekstiä, joka ei ole terminologisesti kiinnostavaa. Käsitteen nimitys ei ole välttämättä edes samassa lauseessa määritelmän kanssa. Tämä hankaloittaa tiedon automaattista hakua korpustyökaluilla, sillä koetin saattaa olla niin kaukana termistä, että termi jää korpusohjelman konkordanssihaun tuottaman tekstikontekstin ulkopuolelle. Esimerkiksi yläkäsite ei välttämättä ole samassa virkkeessä kuin alakäsite, jolloin tieto on liian kaukana automaattisen poiminnan hauehdosta.

Jotkut koettimet tuottavat runsaasti roskaa eli osumia, jotka eivät sisällä toivotunlaista terminologista tietoa. Varsinkin englanninkielisen koettimen *is a* suomalainen vastine *on* on ongelmallinen. Se tuottaa runsaasti roskaa, sillä vain osa väitelauseista, joissa *olla*-verbi on joko muodossa *on* tai *ovat*, sisältää termejä tai käsitetietoa. Koettimien tuottamaa roskaa on yritetty poistaa esimerkiksi tunnistamalla niin sanottuja negatiivisia koettimia eli sanoja tai ilmauksia, jotka eivät koskaan ilmaise semanttisia suhteita (Ahmad & Fulford 1992: 17 lähteessä Bowker 1996b: 36). Roskaa on yritetty poistaa myös tarkentamalla hakua katkaisumerkin käytön sijasta (Christensen 2000: 243–249). Tanskankielisiä koettimia tutkinut Christensen käytti tarkennetussa haussa hakuehtona verbin ja preposition yhdistelmää. Sama keino pätee kaikkiin niihin kieliin, joissa prepositiot kuuluvat kielen systeemiin. Roskaongelmaan ei kuitenkaan toistaiseksi ole yleispätevää ratkaisua.

Koettimien käyttökelpoisuuteen vaikuttaa ratkaisevasti se, kuinka suuren osan hakulausekkeen antamista osumista täytyy olla käyttökelpoisia, jotta hakulauseke voidaan hyväksyä koettimeksi (Grinsted 2000: 48–49). Jos suurin osa osumista on roskaa eli ei sisällä terminologisesti kiinnostavaa tietoa, koetinta ei ehkä kannata käyttää automaattisessa tiedonhaussa. Soininen (1999a: 51) viittaa Cogniterm-ryhmän saamiin tuloksiin, joiden mukaan automaattisessa haussa jäi löytymättä runsaasti mahdollisia rakenteita ja noin puolet automaattisesti haetuista rakenteista oli virheellisiä. Jotta koettimet todella nopeuttaisivat sanastotyötä, olisi hakulausekkeiden listaan pystyttävä löytämään ne koettimet, joista on eniten hyötyä, ja vain ne.

Koettimet eivät paljasta kaikkea terminologisesti kiinnostavaa tietoa tekstistä, koska niitä ei välttämättä aina esiinny termien tai määritelmien yhteydessä. Tästä huolimatta tutkijat ovat olleet yksimielisiä siitä, että niiden tutkimusta kannattaa jatkaa, jos niiden avulla tekstistä saadaan automaattisesti poimittua edes osa termeistä ja käsitetiedosta.

Olen edellä selostanut lyhyesti terminologisia koettimia käsittelevää tutkimusta ja kuvannut koettimien käyttöä käsitetiedon poiminnassa teoreettisesta näkökulmasta. Ennen koetintutkimukseni empiiristä osuutta selostan tutkimusaineistoni koostamisperiaatteita ja tutkimusmenetelmäni.

### 7.3 Terminologisten koettimien tunnistuksessa käytetty aineisto ja menetelmä

Terminologisen tutkimuksen kohde on yleensä tarkoin rajattu ja termistöt ovat alakohdaisia. Tästä syystä terminologioiden täytyy jokaista tutkimusta varten koota eri aineisto, jota ei ehkä koskaan voi käyttää uudelleen ainakaan sellaisenaan (Ahmad & Rogers 2001: 732–733). Kertakäyttöisyys ei kuitenkaan saisi merkitä laadusta tinkimistä. Aineiston koostaminen ei ole yhdentekevää, sillä saadun tiedon laatu riippuu aineiston laadusta (Gamper & Stock 1998/1999: 150).

Tekstikorpuksen tai tekstiarkiston kokoaminen omiin tarpeisiin on nykyisin täysin mahdollista esimerkiksi kääntäjälle tai sanastotyötä tekevälle terminologille. Tekstejä on yhä enemmän saatavilla sähköisesti tallennettuina, jolloin niiden analyysi korpustyökaluilla tai terminpoimintaohjelmilla on mahdollista. Oman korpuksen kokoaminen saattaa olla jopa ainoa vaihtoehto, sillä käyttötarkoitukseen sopivaa korpusta ei ehkä ole olemassa tai esimerkiksi tutkimuslaitoksissa kootut korpuksset eivät tekijänoikeussyistä välttämättä ole ulkopuolisten käytettävissä.

Tämän tutkimuksen aineisto on tallennettu sähköisesti sellaisessa muodossa, että sen käsittely tutkimukseen valituilla menetelmillä on mahdollista. Aineistoa voidaan mielestäni nimittää korpukseksi<sup>1</sup>, mutta lisäksi sitä voidaan nimittää *tekstiarkistoksi* (textual archive; vrt. situational archive; Temmerman 2000: 53), sillä se on tiettyyn tarkoitukseen koottu tekstiaineisto, kun taas korpusta ei välttämättä koota tiettyyn ennalta tarkasti määritettävään tarkoitukseen.

Aineistoni täyttää Ahmadin ja Rogersin (2001: 726) tekstiarkistolle asettamat vaatimukset. Niiden mukaan aineiston on oltava fyysisesti yhdessä paikassa julkaisumuodosta riippumatta. Lisäksi aineisto pitää arvioida, valita ja järjestää systemaattisesti eksplisiittisten kriteerien mukaan. Aineistoani voi sanoa myös *täsmäkorpukseksi* (vrt. Varantola 1999: 131), koska siihen on pyritty valitsemaan tekstejä, joissa todennäköisesti on runsaasti merenkulun turvallisuutta koskevaa käsiteltietoa. Seuraavissa jaksoissa selostan tarkemmin tutkimusaineistona toimivien korpusten koostamisperusteita ja niitä periaatteita, joilla tämän tutkimuksen korpus on koottu.

### 7.3.1 Tekstikorpuksen laajuus

Korpuksen tai tekstiarkiston koostamista koskevassa keskustelussa on pohdittu erityisesti korpuksen laajuutta, joka ilmaistaan yleensä sanoina. Laajuus voi vaihdella erittäin paljon. Laajimpia korpuksia ovat englannin yleiskielen korpuksset, esimerkiksi 100 miljoonan sanan BNC eli British National Corpus, mutta esimerkiksi Helsingin yliopiston kielitieteen laitoksen korpuskokoelmassa olevan somalin kielen korpuksen koko on vain noin 6 000 sanaa. Terminologian tutkimuksissa on käytetty hyvinkin erikokoisia korpuksia. Christensenin (2000: 247, 277–278) koettimien toimintaa käsitelleessä tutkimuksessa aineistona oli 110 000 sanan hydraulikkakorpus, joka hänen mukaansa riittää illustroimaan tutkittavaa ilmiötä, mutta on liian pieni käytännön sanastotyöprojektille. Christensen olikin täydentänyt aineistoaan 6 miljoonan sanan populaaritieteen teksteistä koostuvalla korpuksella. Pearson (1998: 65) käytti tutkimuksessaan kolmea laajuudeltaan huomattavan erilaista korpusta: 4,7 miljoonan sanan erikoisalakorpusta, miljoonan sanan oppikirjakorpusta ja 230 000 sanan asiantuntijoilta asiantuntijoille -korpusta. Pearson päätyi tulokseen (1998: 65), että määritelmätiedon määrä ja ilmaisutapa ei riipu niinkään korpuksen koosta kuin siitä, kuka on tekstin kirjoittaja ja kenelle teksti on kirjoitettu.

Eri tarkoituksiin aiemmin koottujen korpusten koko vaihtelee muutamasta tuhannesta sanasta sataan miljoonaan sanaan, joten korpuksen sanamäärälle on vaikea asettaa yleispätevää ala- tai ylärajaa. Korpuksen on joka tapauksessa oltava riittävän laaja, jotta se on edustava tarkoitukseensa. Puoliautomaattinen käsittely vaatii laajemman korpuksen kuin manuaalinen käsittely. Jos korpus on tarkoitus käsitellä sekä puoliautomaatti-

---

<sup>1</sup> Käännöstieteessä sanalla *korpus* on usein viitattu suhteellisen pieniin tekstikokoelmiin, joita ei ole tallennettu sähköisesti (Baker 1995: 225). ISO-standardin 1087–2:2000(E/F) antaman korpuksen määritelmän mukaan kuitenkin vain sähköisesti tallennetut tekstit muodostavat korpuksen: ”systematic collection of machine-readable texts -- or parts of text prepared, coded and stored according to predefined rules” (ISO 1087–2:2000(E/F): 2)

sesti että manuaalisesti, on korpuksen oltava toisaalta riittävän laaja, jotta tulokset ovat luotettavia, ja toisaalta riittävän pieni, jotta manuaalinen käsittely on mahdollista. Myös aineiston saatavuus asettaa rajoituksia korpuksen koolle. Erityistarkoitukseen kootun korpuksen jonkinlaisena ihannekokona on pidetty miljoonaa sanaa (ks. esim. Pearson 1998: 56). Jos tutkimus keskittyy kapean erikoisalan kieleen tai tiettyyn tekstilajiin, voi korpus olla pienempikin eli esimerkiksi 500 000 sanaa. Koko riippuu lisäksi aineiston saatavuudesta ja käytettävissä olevasta ajasta. Aineiston tulisi olla laaja, mutta suurempi ei aina ole parempi, vaan tärkeämpää on, että korpus sopii siihen tarkoitukseen, johon sitä käytetään (Bowker & Pearson 2002: 9–10, 45). Korpuksen kokoa kasvattamalla ei välttämättä saadakaan parempia tuloksia, sillä esimerkiksi Strehlow (2001: 433) on päätenyt tulokseen, jonka mukaan tietokoneohjelmalla tehdyn terminhaun tehokkuus heikenee kun aineiston koko kasvaa. Jossakin vaiheessa saavutetaan kyllästymisaste eli eri termien määrä ei enää kasva, vaikka korpuksen koko kasvaa. Tämä sääntö pätee ainakin silloin, kun aihealue on suppea. Tulos riippuu myös siitä, kuinka paljon tekstissä on termejä ja käsitetietoa muuhun kieliaineeseen verrattuna. Pearson (1998: 57) esittää, että paras periaate on koota korpus ensin niin laajaksi kuin mahdollista ja kasvattaa sitä myöhemmin, sillä käytännössä korpuksen kokoa rajoittavat materiaalin saatavuus ja tekijänoikeuskysymykset.

Bowkerin ja Pearsonin kokemuksen mukaan hyvin suunniteltu korpus, jonka laajuus on kymmenestä tuhannesta useisiin satoihin tuhansiin sanoihin, on käyttökelpoinen erikoiskielen tutkimuksessa (Bowker & Pearson 2002: 48). Ahmad ja Rogers (2001: 735) sanovat, että koska erikoiskielen teksteissä sanasto on suppeampi kuin yleiskielen teksteissä, riittää aluksi noin 100 000 sanan korpus. Muutaman kymmenen tuhannen sanan tekstikokoelma on jo hyödyllinen, koska erikoiskielessä variaatio on vähäisempää kuin yleiskielessä ja erikoiskielille tyypillisen tiiviin ilmaisutavan ansiosta lyhyessäkin tekstissä voi olla paljon relevanttia kielellistä ainesta (vrt. Ahmad & Rogers 2001: 735–736). Omaan käyttöön laadittavaa sanastoa varten kootun korpuksen koko voi olla esimerkiksi 30 000 sanaa eli 10 noin 3 000 sanan artikkelia (Bowker & Pearson 2002: 144).

Hyvistä periaatteista ja huolellisesta suunnittelusta huolimatta korpus on joka tapauksessa keinotekoinen kokonaisuus (artefact; Pearson 1998: 43) ja aineiston valintaan vaikuttavat käytännölliset rajoitteet ja aineiston saatavuus (Kageura & Umino 1996: 282). Jotkin toivotut lähteet eivät ehkä ole saatavilla, aikakauslehtien numeroita saattaa puuttua, esitelmiä ei ole saatavana kirjallisina tai kirjan painos on loppuunmyyty (Temmerman 2000: 53). On muistettava, että aineiston avulla saadut tulokset ovat vain esimerkkejä, näytteitä tai mahdollisia tilanteita.

Tutkimukseni aineisto on korpukseksi pieni, sillä sen yhteenlaskettu sanamäärä on noin 100 000 sanaa. Sitä voidaan kuitenkin aiempien tutkimusten perusteella pitää riittävänä illustroimaan tutkittavan kapean erikoisalan käsitteistöä ja termistöä. Aineisto on toisaalta pyritty saamaan niin pieneksi, että sen käsittely käytössä olevilla resursseilla olisi mahdollista, ja toisaalta niin suureksi, että se olisi riittävän kattava. Vaarana on, että runsaasti terminologista tietoa sisältäviä tekstejä jää aineiston ulkopuolelle. Koska tutkimus on laadullinen eivätkä tutkimustyöhön osoitetut resurssit anna mahdollisuutta suuren aineiston käsittelyyn, keskityn tutkimuksessani siihen käsitetietoon, joka aineistosta nousee. Tutkimuksen ei ole tarkoituksena antaa kattavaa kuvaa merenkulun turvallisuuskäsitteistä ja -termeistä.

Tärkein korpuksen tai tekstiaineiston koostamiseen vaikuttava seikka on korpuksen käyttötarkoitus. Tiettyyn tarkoitukseen kootun aineiston koostamiseen ei ole kuitenkaan yleisesti päteviä kriteerejä (Pearson 1998: 41). Aineiston valinnan ja järjestämisen tueksi voidaan tosin esittää eräitä yleisiä periaatteita ja jaotteluperusteita. Esimerkiksi Pear-

son (1998: 52–55) luokittelee tekstit ulkoisten ja sisäisten perusteiden mukaan. Ulkoisia ovat ei-kiellelliset perusteet, kuten julkaisumuoto, tekstilaji, viestintätilanne tai tekstin funktio. Ulkoisia perusteita lienevät myös tekstin ikä ja pituus. Sisäisiä ovat esimerkiksi tekstin aihealue, tyyli ja kieli. Olen käyttänyt Pearsonin luokittelua apuna tutkimukseni tekstiarkiston koostamisessa, ja seuraavissa jaksoissa selostankin tekstien valintaa tämän luokittelun pohjalta.

### 7.3.2 Tekstien julkaisumuoto

Tekstin julkaisumuoto on yksi Pearsonin (1998: 52–53) mainitsemista aineiston valinnan ulkoisista perusteista. Julkaisumuodon mukaan aineisto voidaan jakaa kirjalliseen, sähköiseen ja suulliseen aineistoon. Kirjallisia lähteitä ovat sanakirjat, sanastot, tietosanakirjat, oppikirjat, standardit, erikoisalan lehdet, luettelot ym. erikoisalan julkaisut. Sähköisiä lähteitä ovat CD-ROMit ja Internet. Suullisia ovat esimerkiksi erikoisalan asiantuntijat. Pearsonin (2000: 95) mukaan aineistoon valittavien tekstien tulisi olla julkaistuja, koska näissä käytetään ainakin jollakin tasolla hyväksyttyjä termejä ja määritelmiä.

Tämän tutkimuksen tavoite on tuottaa tietoa sellaisten työkalujen kehittämiseksi, jotka auttavat kääntäjää, tulkkia, terminologia, erikoisalan asiantuntijaa, opettajaa tai opiskelijaa poimimaan kirjoitetuista teksteistä tietoa tietyn erikoisalan termeistä ja käsitteistä ja varmistumaan tiettyjen termien nimeämien käsitteiden vastaavuudesta tai mahdollisista käsite-eroista eri kielissä. Poimittuja termejä ja käsitetietoa käytetään mahdollisesti sanastotyössä tai erikoisalaan tutustumisen välineenä. Tämä tavoitteenmäärittely on ohjannut tutkimusaineiston koostamista niin, että lähes kaikkien aineistoon valittujen tekstien alkuperäinen julkaisumuoto on kirjallinen, vaikka tekstit onkin myöhemmin tallennettu sähköisesti. Tekstillä tarkoitan tässä tutkimuksessa pitempää kuin muutaman sanan kokonaisuutta, vaikka joidenkin määritelmien mukaan yksikin sana voi muodostaa tekstin. Esimerkiksi Martin, Al ja van Sterkenberg (1983: 77) määrittelevät tekstin seuraavasti: ”- a set of one or more (written or spoken) words forming a consistent whole from a semantic, syntactic and pragmatic point of view.” Kuten Martinin ym. määritelmästä ilmenee, teksti voi olla kirjallinen tai suullinen. Tutkimukseni aineistossa ei kuitenkaan ole suullisia tekstejä, koska tarkoituksena on tutkia menetelmiä, joilla termejä ja käsitetietoa voi poimia kirjoitetuista erikoisalateksteistä.

### 7.3.3 Tekstilaji

Toinen Pearsonin (1998: 52–53) mainitsema tekstien ulkoinen jaotteluperuste on tekstilaji (genre). Tekstilajista on tehty lukuisia tutkimuksia, mutta sen määrittely on osoittautunut hankalaksi. Esimerkiksi Swales (1990) käyttää määrittelyyn noin kuusi sivua kirjassaan *Genre Analysis*. Bhatia (2005: 17) määrittelee käsitteen lyhyemmin vahvasti konventionaalistuneeksi viestintätapahtumaksi. Tekstilajin käsitteessä näkyy siis tekstin sosiaalinen rooli. Sekaannusta saattaa aiheuttaa toinen tekstien luokitteluun liittyvä käsite, tekstityyppi, joka liittyy tekstien luokitteluun niiden lingvististen ominaisuuksien perusteella. Tekstilajin käsite onkin helpompi hahmottaa esimerkkien kuin määritelmien kautta. Käsitteen juuret ovat kirjallisuustieteessä, jossa romaanit, novellit, näytelmät ja runot muodostavat kukin oman tekstilajinsa. Myöhemmin termiä *tekstilaji* on alettu käyttää myös muiden tekstien yhteydessä. (Yli-Jokipii 2006: 110–112.) Oman tekstilajinsa muodostavat esimerkiksi kaunokirjallisuus, tieteellinen kirjallisuus, sanomalehdet, aikakauslehdet, raportit, säädökset, patentit ja ohjekirjat (Pearson 1998: 52; Yli-Jokipii 2006: 111). Yleisin tekstilaji erikoistarkoitukseen kootuissa korpuksissa on Pearsonin



(1998: 57) mukaan tieteellinen kirjallisuus. Myös erikoisalan aikakauslehtiartikkelit ovat suosittuja, koska ne ovat helposti saatavilla, ilmestyvät säännöllisesti ja niiden sanasto on käytössä olevaa sanastoa. Jaotteluperusteena tekstilaji on kuitenkin ongelmallinen, sillä universaalin tekstilajiluokittelun puuttuessa tutkijat käyttävät omia luokittelujaan, mikä hankaloittaa tulosten vertailua (Pearson 1998: 53).

Koska tutkimukseni tavoite on sanastotyön tukeminen, on tarkoituksenmukaista, että tutkimusaineisto jäljittelee sanastotyössä käytettävää aineistoa. Sanastotyössä yleisesti hyväksytyn periaatteen mukaan sanastotyön lähteitä ovat auktorisoitu aineisto (lait ja asetukset, viranomaisasiakirjat sekä standardit), tiedeyhteisön hyväksymä aineisto (muun muassa tutkimusraportit, oppikirjat, ammattilehdet ja erikoisalan sanakirjat), ajankohtainen vakiintumaton aineisto (muun muassa esitteet, ohjekirjat ja mainokset) ja suulliset lähteet (ks. esim. Sanastotyön käsikirja 1988: 142–143; vrt. myös Ahmad 1994: 2; Griněv 2002: 20–23). Saman periaatteen mukaan aineisto järjestetään hierarkisesti niin, että auktorisoituun aineistoon kuuluvat tekstit ovat lähteinä ensisijaisia.

Olen edellä jaksossa 1.2 muotoillut tutkimukseni tavoitteeksi tuottaa tietoa sellaisten työkalujen kehittämiseksi, joiden avulla aidoista erikoisalateksteistä koostuvasta aineistosta voidaan paikallistaa käsitteet sekä poimia termit ja mahdollinen käsitetieto. Tutkimustavoitteen muotoilusta johtuen aineistosta karsiutuvat erikoisalan sanakirjat ja sanastotyössä niihin rinnastettavat standardit. Tutkimusasetelman olisi voinut järjestää niin, että eri menetelmillä toteutetun terminpoiminnan ja käsitetiedon poiminnan tuloksia olisi voinut verrata sanastojen ja standardien tietoihin, sillä sanakirjojen ja standardien käytöllä sanastotyössä ja terminologisessa tutkimuksessa on etunsa. Standardeihin ja sanastoihin otetut termit ovat käyttöön vakiintuneita ja erikoisalayhteisön hyväksymiä. Citkina (1988: 46) pitääkin termiyn osoituksena sitä, että sana tai ilmaus on otettu erikoisalan sanakirjoihin. Lisäksi standardien ja sanastojen määritelmät ovat asiantuntijoiden tarkkaan harkitsemissa.

Standardeilla ja sanastoilla on kuitenkin puutteita, jotka rajoittavat niiden käyttöä sanastotyössä tai terminologisessa tutkimuksessa. Uusia käsitteitä ei yleensä ole sanakirjoissa ja jos on, niin vastaavuuden varmistamiseen tarvittava määritelmä saattaa puuttua. Soininen (1998: 19) sanoo korpuspohjaisen tutkimuksen osoittaneen, että loppujen lopuksi vain harvat termit löytyvät alan sanastoista. Sanakirjat kuvaavat muutaman vuoden takaista termistöä, eivätkä pysy tieteen kehityksessä ajan tasalla, sillä painettujen sanakirjojen ja standardien kokoaminen ja julkaiseminen vaatii vuosien työn (Muravitskaja & Perebejnos 1987: 83). Joskus ilmaukset ovat vanhentuneita tai käytännön elämälle vieraita. Lisäksi sanastoissa on yleensä vain vähän monisanaisia termejä, lyhenteitä, kontekstietoa ja tietoa termien käyttöyhteyksistä tai käytön laajuudesta (Bowker & Pearson 2002: 15). Myös Kobrin (1989: 8) on havainnut, että sanakirjoissa on yleensä vain yksi- tai kaksisanaisia termejä. Sanastoja ei siis voi käyttää yksinomaisten termien oikeellisuuden mittarina. Sanastojen määritelmät eivät välttämättä riitä käsitteestaavuuden varmistamiseen, sillä sanakirja- ja sanastomääritelmissä realisoituu vain osa käsittepiirteistä. Lisäksi määritelmän laatijalla on aina jokin näkökulma käsitteeseen ja määritelmä on laadittu tiettyä käsitejärjestelmää varten. Myös asiantuntijoiden mielipide-erot vaikuttavat määritelmien muotoiluun. Muravitskaja ja Perebejnos (1987: 83) lisäävät sanakirjojen puutteisiin sen, että todellisen kielenkäytön sijasta ne kuvaavat ihannetilaa ja yksinkertaistavat termijärjestelmää. Erikoisalan sanastoja voi pitää eräänlaisina suljettuina järjestelminä, jotka ovat irti käsitteiden ja termien toimintaympäristöstä. Suomessa erikoisalan sanakirjojen ja standardien käyttöä rajoittaa niiden korkea hinta ja se, että suomeksi ei läheskään kaikilta erikoisaloilta ole julkaistu erikoisalan sanakirjaa.



Vaikka päätin jättää sanakirjat ja standardit aineiston ulkopuolelle, jäi jäljelle vielä joukko erilaisia mahdollisia aineistoon valittavia tekstejä. Standardien ohella auktorisoidua aineistoa ovat esimerkiksi lait, asetukset ja viranomaisasiakirjat. Tämän tutkimuksen alkuvaiheessa tutkimusaineisto koostuikin pääosin merenkulun turvallisuuskysymyksiä koskevasta auktorisoidusta aineistosta. Niistä tärkeimpiä ovat Kansainvälisen merenkulkujärjestön IMO:n sopimukset ja päätöslauselmat, jotka ovat kaikkia jäsenvaltioita sitovia. Merenkulun turvallisuuden kannalta tärkeimpiä kansainvälisiä sopimuksia ovat *the International Convention for the Safety of Life at Sea, 1974 (SOLAS)*, *the Convention on the International Regulations for Preventing Collisions at Sea, 1972 (COLREG)*, *the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978, as amended in 1995 and 1997 (STCW Convention)* ja *the International Management Code for the Safe Operation of Ships and for Pollution Prevention (the ISM Code)*. Kaikki edellä mainitut sopimukset on käännetty muun muassa suomeksi, ja ne ovat osa Suomen lainsäädäntöä.

Vaikka kansainvälisten sopimusten, päätöslauselmien ja muun auktorisoidun aineiston tärkeyttä merenkulun turvallisuustermien ja turvallisuuskäsitteitä koskevan tiedon lähteenä ei voi kiistää, päätin jättää auktorisoidun aineiston tutkimusaineiston ulkopuolelle. Tähän oli kaksi syytä. Ensinnäkin vaikka kansainväliset sopimukset ja muut kansainväliset asiakirjat ovat osa kansallista lainsäädäntöä niin Suomessa kuin Venäjällä, niitä ei ole alun perin kirjoitettu kansallisella kielellä vaan englanniksi. Vastaavat suomen- ja venäjänkieliset asiakirjat ovat aina käännöksiä englannin kielestä. Kielen ilmiöiden tutkimuksessa käännöstekstejä parempia lähteitä ovat autenttiset, kirjoittajan omalla äidinkielellään kirjoittamat tekstit, joissa käännösinterferenssi ei ole vaikuttanut kirjoittajan valintoihin. Toiseksi auktorisoitu aineisto edustaa tämän tutkimuksen näkökulmasta pikemmin lain ja hallinnon kieltä kuin merenkulun kieltä. Säädöksissä ja viranomaisasiakirjoissa on toki runsaasti merenkulun termejä ja käsitetietoa, mutta tekstilajin ja käännöskielen vaikutuksesta niitä ei voi pitää aitoina käyttöteksteinä, jos tällaisen tekstin kriteeri on, että se kuvaa todellista kielenkäyttöä erikoisalaviestinnässä. Näistä syistä tämän tutkimuksen aineistossa ei ole kansainvälisiä sopimuksia, kansallisia lakeja ja asetuksia eikä viranomaisasiakirjoja.

Sanastotyön lähteiden luettelosta ovat näin eri syistä karsiutuneet standardit, sanakirjat, kansainväliset sopimukset, lait, asetukset ja viranomaisasiakirjat sekä suulliset lähteet, joita käsitelin lyhyesti julkaisumuodon yhteydessä jaksossa 7.3.2. Sanastotyön periaatteissa mainitusta aineistosta jäljelle jäävät vielä tiedeyhteisön hyväksymä aineisto sanakirjoja lukuun ottamatta (muun muassa tutkimusraportit, oppikirjat ja ammattilehdet) ja ajankohtainen vakiintumaton aineisto (muun muassa esitteet ja ohjekirjat). Tämän tutkimuksen aineisto on koottu näihin ryhmiin kuuluvista teksteistä. Tällä tavalla karsittu aineisto riittää kuvaamaan erikoisalan kieltä, sillä esimerkiksi Bowkerin (1996b: 45) mukaan edustavassa erikoisalakorpuksessa on tasapuolisesti seuraaviin kolmeen ryhmään kuuluvia tekstejä: ohjaavia tekstejä (instructional texts), kuten oppikirjoja, käsikirjoja ja tietosanakirjoja, tieteellisiä tekstejä (advanced texts), kuten tutkimuksia ja asiantuntijahaastatteluja, sekä yleistajuisia tekstejä (popularized texts), kuten lehtiartikkeleita ja mainoksia. Aineistooni kuuluvat artikkelit on julkaistu lehdissä, joilla on arvostettu asema ilmestymismaan merenkulkuyhteisössä. Suomenkielisistä lehdistä aineistossa on Merenkululaitoksen lehti *Meriväylä* ja venäjänkielisistä lehdistä *Морской флот*, jonka perustajien joukossa on muun muassa venäläisten varustamojen liitto *Союз российских судовладельцев*. Lehdellä on myös pitkät perinteet, sillä se on ilmestynyt vuodesta 1886.

Aineistoni on erikoisalan käyttöteksteistä koottu korpus, jossa termit ovat erikoisalakontekstissa (ks. Grinëv 2002: 20–23). Aitojen erikoisalatekstien käyttöä terminolo-

gisen tutkimuksen aineistona puolustavat myös Dubuc ja Lauriston (1997: 85). Niiden etu on, että ne kuvaavat erikoisalan kieltä sellaisena, kuin alan toimijat sitä käyttävät. Kun termejä ja käsitetietoa poimitaan aidoista käyttöteksteistä, on syytä muistaa, että teksteissä termejä saatetaan käyttää ”väärin” ja käsitetieto on vajavaista. Käyttötekstit kuvaavat erikoisalan käsitteistöä aina kirjoittajan valitsemasta näkökulmasta, joka saattaa poiketa alalla yleisesti hyväksytystä näkemyksestä. Aidoissa teksteissä terminologisen tiedon automatisoinnin ongelmat korostuvat, sillä termivariaatio on yleistä ja käsitetieto on piilossa tekstirakenteissa. Jos kirjoittaja on erikoisalan asiantuntija, joka kirjoittaa toiselle saman alan asiantuntijalle, kirjoittajalla ja vastaanottajalla on paljon yhteistä tietoa, jota ei tekstissä ole ilmaistu. Vajavaisten ja virheellistenkin tekstien käyttöä tutkimusaineistona perustelen sillä, että terminologi tai kääntäjä on usein juuri samassa tilanteessa silloin, kun hän yrittää muodostaa kuvaa erikoisalan käsitteistöä. Kaikki tieto siitä, miten erikoisaloilla on tapana viestiä erikoisalan käsitteistä, auttaa tietojen arvioinnissa. Aitojen tekstien käyttö terminologisessa tutkimuksessa ja sanastotyössä onkin lisääntynyt käytännön syistä ja tietotekniikan kehittymisen myötä. Käytännön syy aitojen tekstien käyttöön on esimerkiksi se, että ennen kuin termit ja käsitetieto päätyvät sanakirjoihin tai sanastoihin, ne kootaan erikoisalan asiantuntijoiden kirjoittamista teksteistä tai asiantuntijoilta suullisesti. Siksi on tärkeää kehittää menetelmiä, joilla saadaan poimittua mahdollisimman paljon esimerkiksi sanastotyön tekijöiden tarvitsemaa tietoa.

### 7.3.4 Viestintätilanne

Kolmas Pearsonin (1998: 52–53) mainitsema ulkoinen tekstien jaotteluperuste on viestintätilanne. Siihen vaikuttaa ennen kaikkea viestintään osallistuvien henkilöiden asiantuntijuus erikoisalalla, jolla viestitään. Yksi viestintään osallistuva on tekstin kirjoittaja ja toinen on tekstin vastaanottaja. Erikoisalan tekstit voidaankin luokitella viiteen luokkaan viestinnän osanottajien asiantuntijuuden mukaan (Bergenholtz 1995: 19):

- 1) asiantuntijalta asiantuntijalle (asiantuntijakieli)
- 2) asiantuntijalta puoliasiantuntijalle (semi-expert; esim. oppikirjat)
- 3) puoliasiantuntijalta puoliasiantuntijalle (ammattislangi ja lähialojen asiantuntijoiden kieli)
- 4) asiantuntijalta maallikolle (kansantajuistetut tekstit)
- 5) puoliasiantuntijalta maallikolle.

Luokittelusta puuttuu kuudes mahdollinen viestintätilanne, jossa viestintä tapahtuu maallikolta maallikolle, koska maallikoiden ei oleteta käyttävän keskenään erikoisalan kieltä. Pearsonin (1998: 36–38) esittämä jaottelu poikkeaa Bergenholtzin jaottelusta siinä, että hän on jättänyt jaottelun ulkopuolelle asiantuntijalta maallikolle tapahtuvan viestinnän. Näin populaaritekstit eivät Pearsonin mukaan kuulu terminologisen tutkimuksen aineistoon.

Bowkerin ja Pearsonin (2002: 19, 51) mielestä terminologisen tutkimuksen aineistoon on syytä ottaa vain sellaisia tekstejä, joiden kirjoittajat ovat erikoisalan asiantuntijoita. Myös Pearsonin (2000: 96) mukaan kirjoittajan pitää olla erikoisalalla tunnustettu henkilö tai yhteisö, jolla on vaadittava asiantuntemus kirjoittaa aiheesta. Tekstejä tulisi olla useita usealta eri kirjoittajalta, jolloin yksilöllinen kielenkäyttö ei vaikuta tulokseen. Suomessa merenkulun turvallisuudesta kirjoittavien asiantuntijoiden joukko on kuitenkin suppea, jolloin aineistoon on vaikea saada monen eri henkilön kirjoittamia tekstejä. Venäjällä merenkulun asiantuntijoiden yhteisö on laajempi ja kirjoittajia on useita.

Tutkimukseni aineistoon kuuluu oppikirjoja, erikoisalan lehtiä, tutkimusraportteja ja esitelmiä, joiden kirjoittajat ovat asiantuntijoita. Tekstit on kirjoitettu toisille asiantuntijoille tai tuleville asiantuntijoille. Kirjoittajien ja vastaanottajien perusteella tekstit kuu-

luvat Bergenholtzin luokittelussa kahteen ensimmäiseen luokkaan eli viestintätilanne on asiantuntijalta asiantuntijalle tai puoliasiantuntijalle. Olen valinnut aineistoon asiantuntijoiden toisille asiantuntijoille kirjoittamia tekstejä, vaikka ne ovat tutkimukseni kannalta ongelmallisia, sillä termien runsaudesta huolimatta niissä ei välttämättä ole runsaasti käsitetietoa. Toisille asiantuntijoille kirjoittavat asiantuntijat voivat nimittäin olettaa, että alan peruskäsitteet ovat lukijoille tuttuja, eivätkä tästä syystä määrittele käsitteitä (ks. Varantola 1999: 137).

Tutkimusaineistooni valittujen tekstien funktio on informatiivinen tai opettava. Informatiivisia ovat aineistoon valituista teksteistä raportit ja aikakauslehtiartikkelit ja opettavia oppikirjat. (vrt. Pearson 1998: 53, 61)

### 7.3.5 Aihealue

Tärkeimpänä tutkimusaineiston suunnitteluperusteena pidetään yleensä aihealuetta. Muun muassa Grinöv (2002: 20) sanoo, että terminologian tutkimusaineistoa ei pitäisi koota satunnaisista teksteistä, koska ne eivät anna erikoisalasta objektiivista kuvaa. Myös Pearson (1998: 41) sanoo, ettei tekstien valinta saa olla sattumanvaraista, vaan tekstien tai tekstinäytteiden tulee olla edustavia ja tyypillisiä tutkittavan erikoiskielen tekstejä. Tutkijat ovat käyttäneet erilaisia menetelmiä sen testaamiseksi, käsitteleekö teksti todella tutkittavaa aihealuetta. Yksi mahdollisuus on valita tekstit yhdessä erikoisalan asiantuntijoiden kanssa (vrt. Dubuc & Lauriston 1997: 85). Toinen mahdollisuus on arvioida tekstin kelpoisuutta erikoisalan termien frekvenssin avulla. Yleensä tekstiä ei oteta edes tarkasteluun, jos siinä on vain yksittäisiä termiesiintymiä (Lahtinen 2000: 77). Myös käsitetietoa kannattaa hakea teksteistä, joissa on paljon termejä (Pearson 1998: 66). Tämä päätelmä saattaa vaikuttaa itsestäänselvyydeltä, mutta sen perusteella aineistosta voi karsia tekstit, joissa ei ole termejä, pelkäämättä käsitetiedon menettämistä.

Metakielen ilmiöiden tutkimiseksi ei tekstien välttämättä tarvitse olla yhdeltä kapealta alalta. Esimerkiksi koettimia voisi tutkia yhtä hyvin minkä alan teksteistä tahansa, mutta koska tutkimukseni tavoite ei ole pelkästään koettimien vaan myös käsitteiden ja termien tutkiminen, on tarkoituksenmukaista koota aineisto teksteistä, jotka käsittelevät valitsemaani aihealuetta.

Aihealueen mukaisessa aineiston valinnassa voi käyttää lähtökohtana olemassa olevia luokitteluja, kuten kansainvälistä kirjastojen UDK-luokitusta tai luokittelua tekstin otsikon perusteella (Pearson 1998: 54). Tutkimukseni tutkimusongelman ja tutkimusalueen rajauksessa päätin ottaa aineistoksi merenkulun turvallisuutta käsittelevät tekstit. Koska maailmanlaajuisesti yhtenäistä luokittelua eri aloilta ei ole, olen valinnut aineistohaun lähtökohdaksi kansainvälisen kirjastojen UDK-luokituksen, jossa merenkulku kuuluu luokkaan 656.6. Tähän luokkaan kuuluvista julkaisuista aineistoon on valittu tekstejä pääasiallisesti otsikon perusteella, sillä esimerkiksi Pearson (1998: 57) toteaa, että erityistarkoitukseen kootussa korpuksessa julkaisun otsikko ja sisällysluettelo todennäköisesti riittävät ilmaisemaan aihealueen. Jos esimerkiksi merenkulun ammattilehdessä artikkelin otsikossa esiintyy sana *turvallisuus*, artikkelin voi hyvällä syyllä odottaa käsittelevän merenkulun turvallisuutta. Ensimmäisessä aineistonvalinnan vaiheessa kiinnitin huomiota myös tekstin kirjoittajaan ja valitsin tekstejä, joiden kirjoittajat ovat erikoisalalla arvostettuja asiantuntijoita. Toisessa vaiheessa tutustuin teksteihin pintapuolisesti ja valitsin tutkimusaineistoksi sellaiset kirjalliset lähteet, joissa on runsaasti terminologisesti kiinnostavaa tietoa. Jätin kuitenkin aineiston ulkopuolelle artikkelit, joissa on käsitelty yksityiskohtaisesti laitteiden teknisiä ratkaisuja ja ominaisuuksia.

sia. Toinen vaihe oli tarpeen, koska tutkimukseni on ennen kaikkea laadullinen. Tällöin jo aineiston valinnassa on arvioitava sen hyödyllisyyttä tutkimuksen tavoitteiden kannalta. Tästä näkökulmasta teksti on hyödyllinen vain, jos siinä on merenkulun turvallisuustermejä ja käsitetietoa merenkulun turvallisuuskäsitteistä.

### 7.3.6 Tyyli

Tyyli on vaikeasti määritettävä tekstin ominaisuus, koska tyyli voidaan käsittää monella tavalla. En pohdi *tyyli*-käsitettä tarkemmin, koska se ei mielestäni ole tarpeen tutkimusaineistoni esittelyn yhteydessä. Totean vain, että tyyli voi tarkoittaa kirjoittajan persoonallista ilmaisutapaa tai kielellisen ilmaisun yleistä lajia, joka on sidoksissa viestinnälliseen ympäristöön ja kieliyhteisön tapaan viestiä. Tutkimusaineistoni yhteydessä tarkoitin tyylillä kielellisen ilmaisun lajia, tyylillistä muotoa, jolla tekstin sisältö ilmaistaan (ks. Korpela 2007; Saukkonen 1984).

Merenkulun kieli ei ole homogeeninen, vaan siihen sisältyy tyylilajiltaan hyvinkin erilaisia alakieliä. Jos merenkulun kieltä kuvataan jatkumolla, sen toiseen päähän sijoittuvat sopimusten ja säädösten muodollinen kieli sekä kielet kielen sisällä, kuten radioviestinnässä käytettävät standardilauseet, ja toiseen päähän merimiesslangi. Tutkimukseni aineisto on valittu tämän jatkumon keskivaiheilta. Muodolliset sopimus- ja säädöstekstit olen jättänyt aineiston ulkopuolelle samoin kuin puhekieliset tekstit (ks. jaksot 7.3.2 ja 7.3.3). Aineistoon valitut suomen- ja venäjänkieliset tekstit ovat asiatekstejä, joiden perustyyli on asiatyyli. Sille ovat tyypillisiä kahden tai kolmen lauseen pituiset virkkeet ja yksinkertainen kielellinen rakenne. Aineistoni tekstit noudattavat asiatekstin tyyliä myös siten, että teksteissä ei käytetä huudahduksia, katkonaisia lauseita, murteellisuuksia, puhekielisyyksiä eikä lukijan suoraa puhuttelua. (Korpela 2007.)

### 7.3.7 Tekstien ikä

Terminologista tutkimusta on arvosteltu siitä, että se keskittyy tarkastelemaan vain nykyhetken tilannetta. Käsitteiden ja termien muuttumisen tutkimus tarjoaisikin mielenkiintoisen näkökulman kognitiivisiin prosesseihin, joiden kautta käsitteen- ja terminmuodostus tapahtuu. Jotta käsitteiden ja termien muutoksen tutkiminen olisi mahdollista, tekstien tulisi olla eri-ikäisiä, koska käsitettä selitetään laajasti yleensä silloin, kun se on uusi, mutta määritelmät ja selitteet harvenevat, kun käsitteestä tulee osa asiantuntijoiden yhteistä tietoa (Bowker & Pearson 2002: 52). Eräs työni tavoite olikin aluksi tutkia käsitteiden ja termien muutosta, ja tämä tavoite lähtökohtana kokosin aineistoa noin 30 vuoden ajalta. Työn tarkoitus kuitenkin muuttui työn edistyessä, ja päädyin lopulta koostamaan tutkimusaineistoni vuosina 1992–2004 julkaistuista lähteistä. Aineisto on siis koottu reilun kymmenen vuoden ajalta, ja aineiston iän perusteella tutkimustani voidaan pitää nykyhetken tilanteen tarkasteluna eli synkronisena tutkimuksena. Vaikka käsitteiden ja termien muutoksen tutkimus ei enää työn edistyessä ollutkaan varsinainen tavoitteeni, oli mielenkiintoista todeta, että näinkin lyhyt ajanjakso riittää käsitteiden ja termien muutoksen tutkimiseen alalla, jolla kehitys on nopeaa ja jolla syntyy uusia käsitteitä ja otetaan käyttöön uusia termejä.

### 7.3.8 Tekstien pituus

Tutkijat eivät ole olleet aivan yksimielisiä siitä, pitäisikö tekstiaineistoon sisällytettävien tekstien olla kokonaisia tekstejä vai riittävätkö tekstinäytteet edustamaan tutkittavaa kieltä. Ahmad ja Rogers (2001: 726) ovat sitä mieltä, että tekstien tulee olla kokonaisia. Myös Bowker ja Pearson (2002: 49) kannattavat tätä näkemystä, koska kiinnostavat käsitteet, termit, ilmaukset ja kontekstit voivat sijaita missä tahansa tekstin osassa (ks. myös Pearson 1998: 59). Strehlow (2001: 433) puolestaan on päätenyt tulokseen, että lyhyistä tiivistelmistä tehty terminhaku on tehokkaampaa kuin pitkistä teksteistä tehty haku. Strehlowin mukaan tämä johtuu siitä, että tiivistelmissä, indeksitermilistoissa ja otsikoissa ollaan tarkempia termien monoseemisyydestä ja mononyymisyydestä kuin kokonaisissa teksteissä, joissa termeillä on tyylin takia paljon variaatiota. Tutkimukseeni valitut tekstit ovat enimmäkseen kokonaisia tekstejä. Niistä on kuitenkin poistettu kaaviot ja taulukot, sillä niitä ei voi tallentaa tutkimuksessa käytettyjen tietokoneohjelmien vaatimaan tekstitiedostomuotoon.

Soinisen (1998: 7) tekemässä terminpoimintakokeessa ilmeni, että pitkissä teksteissä termin frekvenssi ei ole yhtä hyvä termiyden mittari kuin lyhyissä teksteissä. Tulos on jokseenkin yllättävä, sillä voisi olettaa, että mitä pitempi samaa aihetta käsittelevä teksti on, sitä useammin siinä toistuvat termit. Termien ominaisuutena pidetään nimenomaan pitkienkin termien muodon pysyvyyttä, joka ei ole tavanomaista yleiskielen sanojen yhdistelmille, sillä yleiskielen sanat esiintyvät vain harvoin juuri samalla tavalla yhdistyneinä. Termien muodon pysyvyyttä käytetään hyväksi termien esiintymien lukumäärään perustuvissa termintunnistusmenetelmissä.

Pitkien tekstien valintaa puoltaa se, että niissä on useampia termin esiintymiä ja asioista puhutaan perusteellisemmin kuin lyhyissä, jolloin niissä todennäköisemmin on käsiteltäviä. Koska pitkissä teksteissä käsitellään useampaa käsitettä kuin lyhyissä, myös käsitteiden väliset suhteet selviävät todennäköisesti paremmin pitkistä teksteistä. Esimerkiksi Katz (1996: 21) on kuitenkin todennut, että saman terminologisen ilmaisun (phrase) esiintymät eivät enää lisääny, kun asiakirjan pituus ylittää noin 2 750 sanaa. Tutkimusaineistoni tekstien pituus sanaluvulla mitattuna vaihtelee 647 sanasta (hv92) 38 232 sanaan (Bezop). Kolmen tekstin pituus on alle 1 000 sanaa, seitsemän tekstin pituus on 1 000–10 000 sanaa ja neljässä tekstissä on yli 10 000 sanaa. Aineistossani on siis sekä lyhyiksi että pitkiä luokiteltavia tekstejä.

### 7.3.9 Yhteenveto tutkimusaineistosta

Olen edellä jaksoissa 7.3.1–7.3.8 selostanut tutkimusaineiston koostamisperiaatteita silloin, kun tavoitteena on tehdä sanastotyötä tai terminologista tutkimusta suppealta erikoisalalta. Yleisten koostamisperiaatteiden käsittelyn yhteydessä olen esitellyt myös niitä periaatteita, joilla oman tutkimukseni aineisto on koottu. Tässä jaksossa esitän nämä periaatteet kootusti.

Tutkimukseni alkuvaiheessa asetin tavoitteeksi tuottaa tietoa sanastotyön välineiden kehittämiseksi. Tästä syystä pyrin aluksi noudattamaan niitä aineiston koostamisperiaatteita, joita sanastotyössä perinteisesti noudatetaan. Aineistosta karsiutui kuitenkin eri syistä suurin osa sellaisesta aineistosta, jota sanastotyössä suositellaan käytettäväksi (ks. jakso 7.3.3). Aineistoon jäi asiantuntijoiden toisille asiantuntijoille tai tuleville asiantuntijoille kirjoittamia asiatyylisiä, autenttisia tekstejä, joiden funktio on informatiivinen tai opettava.



Tarkasteltavaksi erikoisalaksi olen valinnut merenkulun turvallisuuden ja tarkasteltaviksi kieliksi suomen ja venäjän. Koska tutkimukseni aineisto koostuu merenkulun turvallisuutta käsittelevistä teksteistä, tutkimus tuottaa todennäköisesti listan koettimista, jotka ovat käyttökelpoisia merenkulun kielessä, mutta joiden käyttökelpoisuus muiden alojen kielessä vaatii lisätutkimusta. Kirjoittajakohtaisuuden vaikutusta tutkimuksessa on pyritty vähentämään ottamalla aineistoon tekstejä monelta eri kirjoittajalta. Suomenkielisen aineiston valinnassa tämä oli kuitenkin jokseenkin vaikeaa, sillä suomeksi kirjoittavia merenkulun turvallisuuden asiantuntijoita on varsin vähän. Venäläisen aineiston valinnassa ongelma ei ollut kirjoittajien harvalukuisuus, vaan tekstien saatavuus. Olen pitänyt tutkimukseni tavoitetta ja valittua erikoisalaa lähtökohtana ja valinnut tutkimusaineistooni tyypillisiä merenkulun turvallisuutta käsitteleviä suomen- ja venäjänkielisiä tekstejä, jotka edustavat erilaisia tekstilajeja. En kuitenkaan ole erottanut tekstilajeja korpuksessa, vaan eri tekstilajeihin kuuluvat lähteet ovat samanarvoisia terminologisen tiedon lähteinä.

Tekstien autenttisuus vaatii, että asiantuntijat ovat kirjoittaneet tekstit omalla äidinkiellään eikä tekstien joukossa ole käännöksiä. Tekstien on myös kuvattava todellista kielenkäyttöä erikoisalaviestinnässä. Olen ulottanut todellisen kielenkäytön vaatimuksen koskemaan myös tekstien ikää niin että tekstien asiasisällön on oltava validia tutkimushetkellä. Valitut tekstit on kirjoitettu vuosina 1992–2004. Tänä aikana käsitteet ja termit ovat muuttuneet alan luonnollisen kehityksen seurauksena, mutta terminologisen tiedon poiminnan kannalta vanhimmatkin aineiston tekstit ovat vielä valideja.

Valittu tutkimusmenetelmä on ohjannut tutkimusaineiston koostamista niin, että kaikki aineistoon valitut tekstit ovat kirjoitettuja ja ne on tallennettu sähköisenä. Puoliautomaattisen ja manuaalisen tutkimusmenetelmän yhdistelmä on vaikuttanut myös tekstikorpuksen kokoon. Korpus koostuu enimmäkseen lyhyistä teksteistä, joiden pituus vaihtelee alle tuhannesta sanasta muutamaan tuhanteen sanaan.

Edellä esitettyjen periaatteiden mukaan kootussa ja sähköisesti tallennetussa suomenkielisessä täsmäkorpuksessa on 12 eripituista tekstiä. Niistä viisi on kirjoitettu esitelmiä, joista neljä on julkaistu. Kolme on tutkimusraportteja, joista yksi on aineistossa vain osittain. Kolme on ohjelmajulistuksia ja 10 aikakauslehtiartikkeleita, jotka on yhdistetty samaan tiedostoon. Aikakauslehtiartikkeleita on valittu aineistoon hyvän saatavuuden ja ajantasaisuuden vuoksi, vaikka niissä onkin oletettavasti vain vähän määritelmätietoa, koska kirjoittaja luultavasti olettaa vastaanottajan tuntevan käsitesisällön ilman määritelmää. Suomenkielisen korpuksen laajuus on 74 215 sanaa. Aineistossa ja sanaluvussa on mukana myös terminpoiminnassa käytetty esitelmäteksti *pw02* (ks. jaksso 6.1). Käytän suomenkielisestä koetinaineistosta jäljempänä lyhennettä *fi12*. Venäjänkielisessä täsmäkorpuksessa on yksi aikakauslehtiartikkeli, jolla on kaksi kirjoittajaa, ja osia oppikirjasta, jolla on kolme kirjoittajaa. Kirjassa kuvataan kattavasti merenkulun turvallisuuden eri osa-alueita. Oppikirjan valinta on perusteltua, sillä oletettavasti määritelmätietoa on eniten juuri oppikirjoissa. Ikävä kyllä vastaavaa suomenkielistä merenkulun turvallisuuden oppikirjaa ei tietääkseni ole julkaistu. Venäjänkielisestä oppikirjasta käytän jäljempänä lyhennettä *Bezop*. Venäjänkielisen korpuksen laajuus on yhteensä 41 025 sanaa. Venäjänkielinen korpus on huomattavasti suppeampi kuin suomenkielinen. Siinä on vain kaksi asiakirjaa, ja sen sanamäärä on vain hieman yli puolet suomenkielisen korpuksen sanamäärästä. Vaikka korpuksot eivät ole vertailukelpoisia, ne täyttävät sen tehtävän, jota varten ne on koottu, eli toimivat terminologisen tiedon lähteenä merenkulun turvallisuuden alalla. Aineistoon valittujen tekstien bibliografiset tiedot ovat lähdeluettelossa *Aineslähteet*-otsikon alla. Edellä käsiteltyjen valintaperusteiden yhteenveto on liitteessä 1.



### 7.3.10 Terminologisten koettimien tunnistuksessa käytetty menetelmä

Mahdollisten koettimien tunnistamismenetelmänä on käytetty aineiston manuaalista käsittelyä. Manuaalisen käsittelyn apuna olen käyttänyt korpustyökalua *WordSmith 4.0*, jota käytetään laajalti korpuslingvistiksi tutkimuksessa ja opetuksessa. Ohjelmalla voi tuottaa sanalistoja, konkordansseja ja kollokaatioita tekstimuodossa sähköisesti tallennetusta aineistosta. Valintaan vaikutti myös ohjelman helppokäyttöisyys, jota esimerkiksi Kemppanen (2008: 72) pitää sen tärkeänä etuna. Ohjelman toiminnoista käytän tutkimuksessani konkordanssitoimintoa, jonka avulla voidaan esittää tietyn merkkijonon kaikki esiintymät tietyssä aineistossa. Toiminnon hakukenttään kirjoitettavasta merkkijonosta käytän nimitystä hakuehto. Konkordanssi tuotetaan tavallisesti muodossa, jossa hakuehdon täyttävät esiintymät konteksteineen tulostuvat tietokoneen näytölle alekkain.

Koska aiempien tutkimusten mukaan (vrt. Christensen 2000: 247–276) paras tulos saadaan hakemalla koettimia yhdessä termien tai termiehdokkaiden kanssa, aloitin manuaalisen haun niistä teksteistä, joista suomalaiset ja venäläiset asiantuntijat ja opiskelijat olivat tunnistaneet termit (ks. luku 6). Suomenkielinen teksti *pw02* on 2 366 sanan esitelmä, jonka suomalainen merenkulun asiantuntija on kirjoittanut suomeksi toisille merenkulun asiantuntijoille. Venäjänkielinen teksti, josta käytän lyhennettä *Eskos*, on venäläisten merenkulun asiantuntijoiden kirjoittama 2 793 sanan artikkeli, joka on julkaistu venäjänkielisessä merenkulun aikakauslehdessä. Aloitin koettimien haun näistä teksteistä, koska niissä saatoin rajoittaa tarkastelun tunnistettujen termien konteksteihin. Koettimien tarkastelun yhteydessä tarkoitan kontekstilla vain tekstikontekstia enkä sisällyttä käsitteeseen tekstinulkoista kontekstia, joka oli tarpeen manuaalisen terminpöiminnan tarkastelussa.

Tekstien *pw02* ja *Eskos* manuaalisessa käsittelyssä tulostin ensin tunnistettujen ja poimittujen termien konkordanssit korpustyökalulla. Konkordanssien muodostamista listoista tarkastelin termien ympäristöä, josta hain käsitetietoa. Tämän jälkeen katsoin, onko termin, käsitesuhteen tai käsitepiirteen yhteydessä käytetty jotakin sellaista ilmaisu, joka olisi mahdollisesti terminologinen koetin. Tavoitteeni oli koota ensin näistä terminpoimintakokeen teksteistä kaikki koetinehdokkaat, jotka viittaavat terminologisesti kiinnostavaan tietoon. Seuraavaksi ryhmittelin mahdolliset koettimet ja niiden ilmaiseman tiedon sen mukaan, ilmaisiko koetinehdokas termiä tai muuta käsitteen nimitystä, määritelmää, vieruskäsitettä, synonymiaa, geneeristä suhdetta, koostumussuhdetta vai funktiosuhdetta. Käsitesuhteiden jaottelussa olen noudattanut jaksossa 7.1.1 selostamiani periaatteita.

Halusin myös selvittää roskan määrän eli todeta, kuinka usein koetinehdokkaan kontekstissa ei ollut terminologisesti kiinnostavaa tietoa. Tästä syystä tulostin teksteistä korpustyökalun avulla kaikki havaittujen koetinehdokkaiden konkordanssit. Näiden ympäristöä tarkastelemalla selvitin, kuinka usein koetinehdokas todella ilmaisi terminologista tietoa ja kuinka usein se oli jossakin muussa tehtävässä. Näin sain selville kunkin koetinehdokkaan tuottamien validien tapausten ja roskan välisen suhteen.

Tämän jälkeen laajensin manuaalista hakua korpuksen muihin teksteihin tavoitteenani koota terminologiset koettimet, jotka tuottavat runsaasti valideja tapauksia, mutta vähän roskaa. Laajemmista aineistoista *fi12* ja *Bezop* ei mahdollisten koettimien haussa ollut mahdollisuutta tulostaa tunnettujen termien konkordansseja, joten koetinehdokkaiden haku oli tehtävä täysin manuaalisesti.

Ryhmittelin teksteistä *fi12* ja *Bezop* manuaalisen haun tuloksena saadut mahdolliset koettimet sen mukaan, mitä terminologisesti kiinnostavaa tietoa ne osoittivat ja tein

konkordanssihaut kaikilla niillä koetinedokkailla, jotka tuottivat vähintään neljä validia tulosta. Koetinhaun vaiheet olivat seuraavat:

- 1) Terminpoiminta teksteistä *pw02* ja *Eskos*
- 2) Konkordanssihaut tunnetuilla termeillä teksteistä *pw02* ja *Eskos*
- 3) Tunnettujen termien kontekstin tarkastelu teksteistä *pw02* ja *Eskos*
- 4) Koetinedokkaiden poiminta teksteistä *pw02* ja *Eskos*
- 5) Manuaalinen koetinedokkaiden poiminta aineistoista *fi12* ja *Bezop*
- 6) Konkordanssihaut mahdollisilla koettimilla aineistoista *fi12*, *Eskos* ja *Bezop*.

Kaikki tutkitut terminologiset koetinedokkaat eivät välttämättä ansaitse nimitystä *koetin*. Aiemmissa tutkimuksissa onkin pohdittu, mikä pitäisi olla käyttökelpoisten hakutulosten osuus, jotta koetinedokasta voisi pitää hyödyllisenä. Joissakin tapauksissa hakutulos voi olla satoja esiintymiä, joista vain 15 prosenttia on käyttökelpoisia, tai esiintymiä on vain muutama, mutta ne kaikki ovat käyttökelpoisia. (Grinsted 2000: 48.) Koettimet eivät paljasta kaikkea terminologista tietoa tekstistä, koska niitä ei aina esiinny käsitetiedon yhteydessä. Osa käsitetiedosta jää siis tekstiin piiloon koettimien käytöstä huolimatta. Poimitun käsitetiedon suhdetta kaikkeen tekstin sisältämään käsitetietoon voitaisiin nimittää saanniksi. Käsitetiedon saantia on kuitenkin vaikea mitata, koska yleensä poimija tietää mitä poimintamenetelmä on tuottanut, mutta ei tiedä piilotiedon määrää eli sitä, mitä menetelmän avulla ei ole saatu poimittua. Tässä luvussa käytänkin tutkimieni terminologisten koetinedokkaiden käyttökelpoisuuden yhtenä mittarina saannin sijasta produktiivisuutta, jolla tarkoitan koetinedokkaan tuottamien validien osumien määrää. Toisena käyttökelpoisuuden mittarina käytän tarkkuutta, jonka esitelin tutkimukseni terminpoimintaa käsitelleessä osassa (ks. jakso 6.2). Terminologisten koettimien tuottaman roskan määrä eli terminologisesti kiinnostamattomien tekstikon tekstien osuus kaikista osumista on helpompi arvioida kuin poimimatta jääneen käsitetiedon määrä. Tietyn koettimen tarkkuus on laskettavissa suoraan konkordanssihaun osumalistasta jakamalla validien osumien määrä kaikkien osumien määrällä. Olen käyttänyt osumien määrää ja koettimien tarkkuutta koetinedokkaiden käyttökelpoisuuden arvioinnissa, jossa olen ryhmitellyt koetinedokkaat kolmeen ryhmään: vahvoihin, hyviin ja heikkoihin. Tutkimusaineistossani parhaiksi osoittautuneita koettimia nimitän vahvoiksi koettimiksi. Tähän ryhmään olen ottanut ne koettimet, joista muotoiltu hakuehto tuotti yli 10 osumaa, joista yli puolet on valideja. Hyvän koettimen hakuehto tuotti yli 10 osumaa, mutta niistä on valideja vain 34–50 prosenttia, tai korkeintaan 10 osumaa, mutta niistä yli puolet on valideja. Heikoiksi olen luokitellut koetinedokkaat, joilla vähintään 10 osumasta valideja on alle kolmannes tai korkeintaan 10 osumasta valideja on alle puolet. Tämän ryhmän koetinedokkaat eivät ansaitse koettimen status-ta.

Produktiivisuuden ja tarkkuuden lisäksi koettimen käyttökelpoisuuteen vaikuttaa sen etäisyys lähimpään termiin tekstissä. Koettimen käyttökelpoisuus automaattisessa terminologisen tiedon poiminnassa paranee huomattavasti, jos se liittyy välittömästi tunnettuun termiin, koska tällöin termi on mahdollista sisällyttää hakuehtoon. Olen tarkastellut koettimien käyttökelpoisuutta myös tästä näkökulmasta, koska tutkimukseni tavoitteena on tuottaa tietoa terminologisen tiedon poiminnan automatisoinnin tueksi.

Seuraavaksi selostan suomalaisten ja venäläisten koetinedokkaiden poimintaa aineistoista *fi12*, *Eskos* ja *Bezop*. Olen käyttänyt kursivointia silloin, kun puhun termistä tai käsitteestä. Tarkoitteita en ole kursivoinut. Venäläisten koetinedokkaiden ja termien suomalaiset vastineet olen antanut suluissa. Joissakin tapauksissa suluissa on venäläinen termi ja suomalainen vastine on tekstissä ennen sulkuja.

## 7.4 Suomalaiset terminologiset koettimet

Aloitin suomalaisten terminologisten koettimien tutkimisen tekstistä *pw02*, josta suomalaiset asiantuntijat olivat poimineet termit. Koettimien poiminnan ensimmäisen vaiheen tavoitteena oli koota alustava suomalaisten koettimien luettelo. Tein tekstistä *pw02* konkordanssihaun tunnetuilla termeillä, ja termien lähikontekstin tarkastelu tuotti koetinehdokkaita, jotka osoittavat termiä, määritelmää, synonymiaa, vieruskäsitteitä, geneeristä suhdetta, koostumussuhdetta tai funktiosuhdetta. Varsin lyhyessä tekstissä on siis tapauksia, joissa koettimet osoittavat erilaista terminologisesti kiinnostavaa tietoa. Tekstin lyhyiden vuoksi yhden koetinehdokkaan tuottamien tapausten määrä on alhainen, usein vain yksi tapaus ehdokasta kohti, eikä niiden perusteella voi tästä syystä tehdä päätelmiä koetinehdokkaan käyttökelpoisuudesta. Jatkoinkin suomalaisten koettimien tutkimista ja koettimien listan täydentämistä laajemmalla suomenkielisellä aineistolla *fi12*, josta ehdokkaiden manuaalinen haku tuotti lisää tapauksia ja myös uusia ehdokkaita. Näillä tein konkordanssihaun, jos tapauksia oli vähintään neljä. Asetin rajaksi neljä validia tapausta, koska mielestäni tämä määrä riittää osoittamaan, että koetinehdokkaan käyttö terminologisen tiedon yhteydessä ei ole satunnaista. Seuraavaksi kuvaan haun tuottamia suomalaisia koetinehdokkaita, jotka olen ryhmitellyt niiden ilmaiseman tiedon mukaan.

### 7.4.1 Käsitteen nimityksen suomalaiset koettimet

Teksteissä nimetään käsitteitä erilaisilla ilmauksilla, joista terminologisesti kiinnostavimpia ovat termit. Käsitteiden ja mahdollisen käsitetiedon paikallistamiseksi kiinnostavia ovat myös vakiintumattomat ilmaukset, jotka eivät vielä ole saavuttaneet termin statusta. Kirjoittajat saattavat nimetä käsitteen myös muulla kielellä kuin millä teksti on kirjoitettu. Varsinkin suomalaiset erikoisalan tekstien kirjoittajat lisäävät tekstiin usein englanninkielisiä termejä. Puhunkin tässä yhteydessä termin sijasta käsitteen nimityksestä, sillä termeihin keskittyminen rajoittaisi koettimien avulla saatavan terminologisesti kiinnostavan tiedon määrää (ks. myös jakso 2.3).

Tekstistä *pw02* tehty koetinhaku tuotti viisi sellaista koetinehdokasta, jotka osoittavat käsitteen nimitystä (ks. myös Pasanen 2005a). Käsitettä nimeävän ilmauksen mahdollisia substantiivikoettimia ovat sanat *termi*, *käsite* ja *ilmaisu* ja mahdollisia välimerkkikoettimia lainausmerkit ja sulkeet. Lainausmerkit tuottivat kuitenkin kelvollisten tapausten lisäksi runsaasti roskaa. Lainausmerkkien ensisijaisena tehtävänä ei ilmeisesti ole toimia käsitettä nimeävän ilmaisun koettimena. Manuaalinen koetinhaku laajemmasta aineistosta *fi12* kasvatti mahdollisten substantiivikoettimien listaa sanalla *nimitys* sekä tuotti lisäksi yhden mahdollisen verbikoettimen, verbin *kutsua*, yhden mahdollisen lyhennekoettimen, lyhenteen *ns.* sekä sen avatun muodon *niin sanottu*. Manuaalisesti poimittujen koetinehdokkaiden testaamiseksi tein näillä ehdokkailla konkordanssihaut aineistosta *fi12*. Konkordanssihaun tulokset olen koonnut taulukkoon 39.

TAULUKKO 39. Käsitettä nimeävän ilmauksen terminologiset koettimet suomenkielissä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
käsite	<i>*käsite/*käsitte/*käsitettä</i>	29	29	100	vahva
termi	<i>*termi* ei: *terminaal*</i>	9	7	78	hyvä
nimi/nimitys	<i>nimi/*nimity*/nimellä</i>	11	7	64	vahva
kutsua	<i>voidaan kutsua/kutsutaan/kutsuttu</i>	8	8	100	hyvä
niin sanottu	<i>niin sanot*</i>	4	4	100	hyvä
ns.	<i>ns.</i>	42	30	71	vahva
sulkeet	<i>(*</i>	821	124	15	heikko
lainausmerkit	<i>”*</i>	127	45	35	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

Konkordanssihaun tavoitteena oli selvittää, tuottavatko tarkasteltavat koetinehdokkaat käsitettä nimeäviä ilmauksia ja kuinka suuri osa osumista ei sisällä tällaista tietoa. Tein konkordanssihaun ensin mahdollisilla substantiivikoettimilla *käsite*, *termi*, *ilmaisu* ja *nimitys*.

**Koetinehdokas *käsite*.** Aineistosta *fi12* tehdyn manuaalisen koetinhaun tuottamien tapauksen perusteella asetin konkordanssihaussa koetinehdokkaan *käsite* hakuehdoksi *\*käsite/\*käsitte/\*käsitettä*. Katkaisumerkki hakusanojen alussa on ilmeisen tarpeellinen, sillä joissakin manuaalisen haun tuottamissa tapauksissa käsitteen nimitys muodostaa koetinehdokkaan kanssa yhdyssanan yhdysmerkin kanssa tai ilman sitä, kuten esimerkiksi sanassa *linjaluotsikäsite*. Konkordanssihaku tuotti 29 osumaa, joissa kaikissa ehdokas toimii käsitettä nimeävän ilmauksen koettimena. Konkordanssilistaan tulostuneiden osumien perusteella koetinehdokas *käsite* ei ole kirjoittajakohtainen, sillä osumat ovat kymmenestä eri lähteestä. Useammin kuin joka toisessa tapauksessa termi tai muu käsitettä nimeävä ilmaus on tekstissä välittömästi ennen koetinehdokasta, kuten esimerkiksi 3 termi *meriturvallisuus*:

- (3) Olen useimmissa meriturvallisuutta koskevissa esityksissäni jakanut meriturvallisuus käsitteen alusturvallisuuteen, väyläturvallisuuteen ja meripelastukseen. (hv98: 1)

Termin ja koetinehdokkaan ortografisessa liittämisessä on kuitenkin vaihtelua. Seitsemässä tapauksessa termin ja koettimen välissä on normin vastaisesti sanaväli, kuten yllä esimerkissä 3, neljässä tapauksessa termi tai termin osa ja koetinehdokas muodostavat yhdyssanan ilman sanaväliä tai yhdysmerkkiä, esimerkiksi *merenkulun turvallisuuskäsite*, viidessä tapauksessa välissä on yhdysmerkki, esimerkiksi *VTMIS-käsite*. Neljässä tapauksessa termin ja koetinehdokkaan välissä on 1–7 sanaa:

- (4) Kommentosiltatyö on toiminnallinen käsite, joka ei rajaa tarkastelua fyysiseen komentosiltaan (kiv97: 2)

Esimerkissä 4 termin *komentosiltatyö* ja koetinehdokkaan *käsite* välissä on kaksi sanaa, mutta kahdessa tapauksessa termi on tekstissä niin paljon ennen koetinehdokasta, ettei se ole tulostunut konkordanssilistaan, ja yhdessä esimerkissä termi on korvattu prono-

minilla. Vain neljäsosassa eli seitsemässä tapauksessa koetinehdokas *käsite* on tekstissä ennen termiä, esimerkiksi *käsite VTM*.

Seitsemässä tapauksessa käsitteen nimityksenä on englanninkielinen termi tai englanninkielinen termi suomenkielisen termin ohella:

- (5) Komentosiltayhteistyön eli Bridge Resource Managementin (BRM) käsite on tullut Suomenkin kauppamerenkulkuun liikenneilmailun kautta. (kiv97: 63)

Esimerkissä 5 on suomenkielisen termin *komentosiltayhteistyö* ohella mainittu käsitettä nimeävä englanninkielinen termi *Bridge Resource Management* ja suluissa englanninkielinen lyhennetermi *BRM*, jota käytetään terminä myös suomen kielessä.

Kokeilin konkordanssihakua myös hakuehdolla *käsite\*/käsitte\*/käsittei\**, joka eroaa aiemmasta hakuehdosta siten, että hakusanan *käsite* lopussa on katkaisumerkki, jotta myös sanan partitiivimuoto *käsitettä* sopisi hakuehtoon. Tämä hakuehto tuotti kuitenkin 20 sellaista osumaa, joissa hakuehdon täyttää jokin *käsitellä*-verbin taivutusmuodoista. Katkaisumerkin lisääminen sanan *käsite* loppuun tuotti siis runsaasti roskaa, mutta ei yhtään uutta validia osumaa.

**Koetinehdokas termi.** Toinen mahdollinen käsitettä nimeävän ilmauksen substantiivikoetin on sana *termi*. Hakuehdolla *termi\** tehty konkordanssihakua aineistosta *fi12* tuotti 17 osumaa, mutta vain viisi näistä oli toivottuja sanan *termi* ja sitä seuraavan termin yhdistelmiä. Peräti 12 osumaa sisälsi sanan *terminaali*. Rajoittamalla hakua sulkulausekkeella *terminaal\** osumalistalle tulostuvat vain sanan *termi* taivutusmuodot, joissa kaikissa koetinehdokas toimii termin koettimena. Jotta hakutulokseen saadaan nekin osumat, joissa termi muodostaa sanan *termi* kanssa yhdyssanan joko yhdysmerkin kanssa tai ilman, on myös hakuehdon alkuun lisättävä katkaisumerkki. Hakuehto *\*termi\** ja sulkulauseke *terminaal\** tuottivat 25 osumaa, mutta nyt konkordanssilistassa oli 14 osumaa, joissa hakuehdon täyttää sana *öljyterminaali*. Jotta *öljyterminaali* ja muut terminaalien alakäsitteet eivät tulostu konkordanssilistaan, on sulkulausekkeen alkuun lisättävä katkaisumerkki. Hakuehto *\*termi\** ja sulkulauseke *\*terminaal\** tuottavat yhdeksän osumaa, joista seitsemässä sana *termi* toimii käsitettä nimeävän ilmauksen koettimena, jolloin koetinehdokkaan tarkkuus on 78 prosenttia. Kahdessa tapauksessa termi muodostaa koetinehdokkaan kanssa yhdyssanan, esimerkiksi *luotsaustermi*, ja kahdessa suomenkielinen termi on tekstissä välittömästi koettimen jälkeen, esimerkiksi *termi vakiorutiinit*. Kahdessa tapauksessa englanninkielinen termi on välittömästi tai yhden sanan päässä koetinehdokkaan jälkeen:

- (6) Englanninkielissä maissa vastaava termi on Standard Operational Procedures (SOP). (kl98: 72)

Esimerkissä 6 koetinehdokkaan *termi* ja englanninkielisen termin *Standard Operational Procedures* välissä on yksi sana, joka on *olla*-verbin muoto *on*. Yhdessä tapauksessa termi on mainittu tekstissä kolme sanaa ennen koetinta, mutta suurimmassa osassa tapauksia käsitettä nimeävä termi ja koetinehdokas ovat tekstissä peräkkäin. Koetinehdokkaan *termi* tuottamat seitsemän validia tapausta ovat neljästä eri lähteestä, joten koetinehdokas ei ole kirjoittajakohtainen.

**Koetinehdokas ilmaisu.** Mahdollisista käsitteeseen nimitystä osoittavista substantiivikoettimista kolmas, sana *ilmaisu*, tuotti aineistosta *fi12* vain vähän osumia. Hakuehto *ilmaisu\** tuotti kolme osumaa, joista kaksi on toivottuja koetinehdokkaan ja käsitettä nimeävän ilmauksen konteksteja:

(7) Tätä kuvataan ilmaisulla horisontaalinen integraatio, - - (pw02: 8)

(8) Tätä kuvataan ilmaisulla vertikaalinen integraatio - - (pw02: 8)

Kumpikin esimerkki on samasta lähteestä, joten koetinehdokas *ilmaisu* on todennäköisesti kirjoittajakohtainen.

**Koetinehdokas *nimitys*.** Substantiivien *käsite* ja *termi* lisäksi substantiivi *nimitys* on aineistosta *fi12* tehdyn manuaalisen koetinhaun perusteella mahdollinen käsitettä nimeävän ilmauksen koetin. Samasta aineistosta tehty konkordanssihaku hakuehdolla *nimi/\*nimitys\*/nimellä* tuotti 11 osumaa, joista seitsemässä koetinehdokas toimii käsitettä nimeävän ilmauksen koettimena. Kolmessa kelpaamattomassa osumassa sana *nimi* tai *nimitys* on yleiskielen merkityksessä, kuten esimerkiksi sanaliitossa *aluksen nimi ja tunnus*. Neljässä validissa tapauksessa seitsemästä termi on välittömästi ennen koetinehdokasta, kuten termi *komentosilta* esimerkissä 9:

(9) Komentosillan nimi periytyy ajalta, jolloin komentosilta oli avoin tila aluksen peräosassa ja se oli ulkonäöltään sillan muotoinen. (kiv97: 41)

(10) Autopilotti-nimityskin tulisi itse asiassa säästää laitteelle, joka todella ohjaa laivaa esteet kiertäen. (js94: 5)

Esimerkissä 10 termi *autopilotti* muodostaa yhdyssanan koetinehdokkaan *nimitys* kanssa. Koetinehdokas toimii siis hyvin samalla tavalla kuin koetinehdokkaat *käsite* ja *termi*, jotka nekin usein muodostavat termin kanssa yhdyssanan. Seitsemän validia tapausta ovat neljästä eri lähteestä, joten koetinehdokas *nimi/nimitys* ei ole kirjoittajakohtainen.

**Koetinehdokas *ns*.** Manuaalinen koetinhaku aineistosta *fi12* osoitti, että lyhenne *ns*. on mahdollinen käsitettä nimeävän ilmauksen koetin. Samasta aineistosta tehdyssä konkordanssihaussa lyhenne *ns*. osoittautuikin produktiiviseksi, sillä hakuehto *ns*. tuotti 42 osumaa, joista 30 on valideja. Kaikissa valideissa tapauksissa merenkulun turvallisuuden termi tai muu käsitteen nimitys on tekstissä heti koetinehdokkaan jälkeen, kuten esimerkissä 11 termi *liiketilänäyttö*:

(11) Esimerkki hyvästä graafisesta näytöstä on ns. liiketilänäyttö. (kiv97: 55)

Lyhenne *ns*. ilmaisee jonkinlaista epävarmuutta sitä seuraavan termin hyväksyttävyydestä tai vakiintuneisuudesta. Yhdessä tapauksessa on epävarmuutta korostettu laittamalla käsitteen nimitys lainausmerkkeihin:

(12) Meriliikennettä ohjataan yhä enemmän ilmailualan tapaan, mikä poistaa ns. ”merten vapauden”, mutta sen sijaan lisää oleellisesti merten turvallisuutta. (sms 02: 32)

Lainausmerkeissä oleva ilmaus *merten vapaus* on kyllä vakiintunut, mutta termin statussa sillä tuskin on.

**Koetinehdokas *niin sanottu*.** Lyhenteen *ns*. avattu muoto *niin sanottu* ei tämän tutkimuksen aineistossa ollut yhtä produktiivinen kuin lyhenne *ns*., mutta tuotti kuitenkin muutaman validin tapauksen. Hakuehto *niin sanot\** tuotti neljä osumaa, joissa kaikissa



termi on välittömästi koetinehdokkaan jälkeen, kuten *VTMIS-järjestelmä* esimerkissä 13 ja *mukavuuslippuvaltio* esimerkissä 14:

- (13) Suomenlahdelle perustetaan niin sanottu VTMIS-järjestelmä. (smr02: 69)
- (14) Linjos-Maunula huomauttaa, että kanaalin kokemusten mukaan myös niin sanotut mukavuuslippuvaltiot ovat tarkkoja omien alustensa käyttäytymisestä. (mv9: 10)

Esimerkissä 14 koetinehdokas on monikossa, kuten myös termi, johon koetinehdokas viittaa.

**Koetinehdokas *kutsua*.** Manuaalisen koetinhaun perusteella myös verbi *kutsua* saattaa olla käsitteen nimityksen koetin. Manuaalinen haku aineistosta *fil2* tuotti nimittäin kuusi kontekstia, joissa koetinehdokas on muodossa *voidaan kutsua*, *kutsutaan* tai *kutsuttu*. Konkordanssihaku samasta aineistosta hakuehdolla *voidaan kutsua/kutsutaan/kutsuttu* tuotti kahdeksan osumaa, joissa kaikissa koetinehdokas viittaa merenkulun turvallisuus-käsitteen nimitykseen:

- (15) Integroitua hallintalaitetta kutsutaan yleisesti joystick-ohjaukseksi. (kiv97: 49)

Esimerkissä 15 koetinehdokas *kutsutaan* viittaa termiin *joystick-ohjaus*, jonka määritelmä, *integroitu hallintalaite*, on annettu ennen koetinehdokasta. Kuudessa tapauksessa termi on koetinehdokkaan jälkeen, joko välittömästi tai yhden sanan päässä, niin kuin esimerkissä 16. Termi on ennen koetinehdokasta kahdessa tapauksessa, joista toisessa termin ja koetinehdokkaan välissä on poikkeuksellisesti peräti kahdeksan sanaa. Rakenteessa on tyypillisesti termin synonyymi, kuten esimerkissä 16, tai yläkäsitteen nimitys, kuten esimerkissä 17:

- (16) Rakenteellista turvallisuutta voidaan kutsua myös passiiviseksi turvallisuudeksi. (hv92: 13)
- (17) Perämeren ja Selkämeren talviliikenteeseen on kehittynyt 1960-luvulta alkaen järjestelmä, jota voidaan kutsua suomalais-ruotsalaiseksi talviliikennejärjestelmäksi. (smr02: 39)

Esimerkissä 17 kontekstissa ennen koetinehdokasta on sana *järjestelmä*, joka on termin *suomalais-ruotsalainen talviliikennejärjestelmä* geneerinen yläkäsite, ja esimerkissä 16 ennen koetinehdokasta on termin *passiivinen turvallisuus* synonyymi *rakenteellinen turvallisuus*. Kolmessa tapauksessa *kutsua*-verbiin liittyy modaaliapuverbi *voida*, kuten esimerkeissä 16 ja 17. Tapauksia on kuitenkin niin vähän, ettei havaintoja voi yleistää muihin aineistoihin päteviksi.

**Koetinehdokas *sulkeet*.** Väli-merkeistä sulkeet ovat konkordanssihaun ja manuaalisen koetinhaun tulosten perusteella mahdollisesti hyvä käsitteen nimityksen koetin. Hakuehto (\* tuotti 821 osumaa, joista 124 sisältää suomenkielisen termin ja sulkeiden sisään sijoitetun, yleensä englanninkielisen termin pitkän muodon tai englannin kielen sanoista muodostetun lyhenteen. Lähes kaikissa tapauksissa termi tai muu käsitteen nimitys on välittömästi ennen sulkeita, kuten esimerkissä 18 termi *väsymys*:

- (18) Väsymys (fatigue) on tekijä, joka usein mainitaan merionnettomuuksien yhteydessä. (kiv97: 77)

Vain kuudessa tapauksessa termin ja vasemman sulkeen välissä on yksi tai kaksi sanaa, kuten esimerkissä 19, jossa termiä *liikenneohjausjärjestelmä* seuraa sana *käyttöönotto*:

- (19) Se suositteli liikenneohjausjärjestelmän käyttöönottoa (Vessel Traffic Service, VTS).  
(kiv97: 85)

Muissa kuin valideissa tapauksissa sulkeissa annetaan lähdeviite, kuten lakikohta tai asiakirjan nimi, tai kalvon tai kappaleen numero. Usein sulkeissa on tekstiin liittyvä lisäys tai selitys, kuten prosenttiluku. Sulkeet ovat kiistatta produktiivinen koetin, mutta hakuehto tuottaa myös erittäin runsaasti osumia, joissa ei ole terminologisesti kiinnostavaa tietoa. Tästä syystä sulkeet ovat käyttökelpoinen koetin vain, jos konkordanssihakua voidaan tehdä yhdessä tunnetun termin tai terminosan kanssa.

**Koetinehdokas lainausmerkit.** Toisen mahdollisen välimerkkikoettimen, lainausmerkkien, toimintaa koettimena tutkin hakuehdolla ”\*”. Se tuottikin 127 osumaa, mutta vain 45 osumassa lainausmerkkien sisällä on merenkulun turvallisuuskäsitteen nimitys. Näissäkin osumissa ilmaukseen sisältyy aina värittyneisyyttä tai epävarmuutta ilmauksen statuksesta, kuten esimerkissä 20, jossa lainausmerkeissä on ilmailusta lainattu mutta merenkulussa vakiintumaton käsite *musta laatikko*:

- (20) VDR-järjestelmät eli ”mustat laatikot” olisi tehtävä pakollisiksi - - (smr02: 89)

Kelvollisten tapausten lisäksi hakuehto ”\*” tuotti 37 kontekstia, joissa kirjoittaja esittää suoran lainauksen, 16 kontekstia, joissa lainausmerkkien sisällä on erisnimi, esimerkiksi asiakirjan tai projektin nimi, ja 29 kontekstia, joissa tavalla tai toisella värittynyt ilmaus on erotettu muusta tekstistä lainausmerkein. Kelvollisten tapausten poiminta ei kuitenkaan ole aivan niin toivotonta, kuin ensin näyttää, sillä kelvollisissa tapauksissa termiin liittyy lainausmerkkien lisäksi aina myös jokin muu koetin, kuten substantiivi *ryhmä*, *luokka*, *käsite* tai konjunktio *eli*. Osumia ryhmittelemällä tai manuaalisesti voidaan hakuehdon ”\*” tuottamien osumien listasta poimia ne osumat, joissa on koetinehdokasta vahvistava toinen koetinehdokas. Karsinnan jälkeen jäljelle jäävät osumat ovat todennäköisesti valideja.

**Yhteenveto.** Tarkastelluista käsitteen nimityksen koetinehdokkaista produktiivisimpia olivat lyhenne *ns.* ja substantiivi *käsite*. Ne ovat vahvoja koettimia, sillä produktiivisuuden lisäksi niillä on muitakin hyviä ominaisuuksia: ne tuottavat vähän roskaa, eivät ole kirjoittajakohkaisia ja soveltuvat puoliautomaattiseen hakuun, koska termi on useimmiten välittömästi ennen koetinta tai välittömästi sen jälkeen. Substantiivit *termi* ja *nimitys* olivat kohtalaisen produktiivisia mutta tuottivat myös roskaa. Vähän roskaa tuottavia koettimia ovat verbi *kutsua* ja sanaliitto *niin sanottu*, mutta osumien vähäisyyden vuoksi ne eivät ole vahvoja koettimia.

#### 7.4.2 Synonymian suomalaiset koettimet

Kansainvälisen ISO-standardin mukaan synonymia on samaa käsitettä nimeävien termien välinen suhde (ks. jakso 5.4). Yleensä synonymialla tarkoitetaan ilmiötä, jossa yhdellä käsitteellä on yhdessä kielessä monta leksikaaliselta muodoltaan erilaista nimitystä. Synonymia ei siis ulotu kielirajojen yli. Kirjoittajat saattavat kuitenkin nimetä käsitteitä myös ilmauksilla, jotka ovat termejä jossakin muussa kielessä kuin siinä kielessä, jolla

teksti on kirjoitettu. Samankielisten synonyymisten termien lisäksi pidän synonyymeinä myös erikielisiä, samaa käsitettä nimeäviä termejä. Näitä termejä kutsun vastineiksi.

Suomenkielisellä esitelmätekstillä *pw02* tehdyssä manuaalisessa koetinhausassa hyviksi synonymian koettimiksi osoittautuivat sulkeet ja konjunktio *eli* (ks. myös Pasanen 2005a: 295). Konkordanssihaun tulokset on esitetty taulukossa 40.

TAULUKKO 40. Synonymian terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit osumat		Arviointi <sup>a</sup>
			f	%	
sulkeet	(*	821	207	25	heikko
	(*)	200	78	39	hyvä
eli	<i>eli</i>	75	27	36	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *sulkeet*.** Esitelmätekstistä tehty konkordanssihaku hakuehdolla (\* tuotti 22 osumaa, joista 19 on kelvollisia termisynonymian konteksteja. Kaikissa valideissa konteksteissa yhtä lukuun ottamatta tekstissä on asiantuntijoiden valitsema termi ennen sulkeita ja sulkeissa samankielinen synonyymi tai erikielinen vastine, jonka asiantuntijat myös valitsivat termilistaansa. Esimerkissä 21 välittömästi ennen sulkeita on asiantuntijoiden valitsema suomenkielinen termi *Euroopan ympäristörannikkovartiosto* ja sulkeissa asiantuntijoiden valitsema englanninkielinen termi *European Environmental Coastguard*:

- (21) Tämä näkyy Euroopan ympäristörannikkovartioston (European Environmental Coastguard) perustamista kohtaan esiintyneistä reaktioista. (pw02: 7)

Konkordanssihaku tekstistä *pw02* tuotti kaikkiaan viisi esimerkin 21 kaltaista osumaa, joissa suomenkielistä termiä seuraa sulkeissa oleva englanninkielinen vastine. Seitsemässä osumassa suomenkielisen termin jälkeen sulkeissa on englanninkielinen lyhenne-termi, joka on viidessä tapauksessa lyhenne *VTS*:

- (22) VTMS-käsite on syntynyt meriliikenteen ohjausjärjestelmistä (VTS) saaduista kokeuksista. (pw02: 4)

Sulkeissa olevaa lyhennetermiä *VTS* voidaan pitää synonyymina suomenkieliselle termille *meriliikenteen ohjausjärjestelmä*, sillä lyhennettä käytetään yleisesti suomenkielisessä tekstissä suomenkielisen termin sijasta. Sulkeiden lisäksi esimerkivirkkeessä on käsitettä nimeävän ilmauksen koetin *käsite*. Koettimet esiintyvätkin usein lähekkäin tekstissä.

Kaikissa valideissa osumissa on sulkeissa vähintään yksi englanninkielinen termi tai lyhennetermi, mutta kolmessa tapauksessa on suluissa englanninkielisen termin pitkän muodon lisäksi englanninkielinen lyhenne tai suomalainen vastine:

- (23) Lisätuloksia saatiin EURET'in RTIS- (Regional Traffic Information Service, Alueellisen liikenteen informaatiopalvelu) ja TAIE-projekteista. (pw02: 5)

Esimerkissä 23 välittömästi ennen sulkeita on englanninkielinen lyhennetermi *RTIS*, jonka asiantuntijat valitsivat termilistaan. Sulkeissa on englanninkielisen termin pitkä muoto *Regional Traffic Information Service* ja suomenkielinen vastine *alueellisen liikenteen informaatiopalvelu*, jotka asiantuntijat niin ikään valitsivat termeiksi. Hakuehdolla (\* suomenkielisestä esitelmästä pystytään siis poimimaan 19 englanninkielistä termiä, jotka johdattavat myös niiden suomenkielisten vastineiden jäljille.

Muiden tutkimusaineistoon valittujen tekstien kirjoittajat käyttävät sulkeita suhteessa vähintään yhtä ahkerasti kuin terminpoimintakokeen tekstin kirjoittaja, sillä konkordanssihaku hakuehdolla (\* tuotti 821 osumaa noin 30 kertaa laajemmasta aineistosta *fi12*. Näin suuren osumamäärän manuaalinen käsittely on työlästä, mutta osumalistaa kannattaa tutkia, sillä jos kaikki kirjoittajat käyttävät sulkeita synonymian tai termivastineiden osoittamiseen yhtä usein kuin tekstin *pw02* kirjoittaja, osumalistassa tulisi olla noin 500 synonymian tai termivastaavuuden tapausta. Nopeasti osumalistaa silmäilemällä voi kuitenkin heti havaita, että näin ei ole, sillä useimmiten sulkeissa on terminologisesti kiinnostamatonta metatekstiä, kuten numeroita, lähdeviitteitä tai viittauksia kuvaan tai kalvoon. Esitelmätekstin *pw02* kirjoittaja on siis käyttänyt sulkeita synonyymisten termien tai vastineiden ympärillä suhteessa paljon useammin kuin muun aineiston kirjoittajat. Kuitenkin myös aineistosta *fi12* tuotetussa konkordanssilistassa on valideja osumia, joita tarkastelen seuraavaksi.

Hakuehdon (\* tuottamista valideista osumista 11:ssä on tekstissä ensin suomenkielinen termi ja sulkeissa sen suomenkielinen synonyymi, kuten esimerkissä 24:

- (24) Myös VTS-järjestelmän (Vessel Traffic Service; alusten ohjaus- ja tukipalvelu) ja luotausohjeiden kehittämistä kaivataan. (sms02: 30)

Paljon useammin tekstissä kuitenkin on välittömästi ennen sulkeita suomenkielinen termi ja sulkeissa englanninkielinen termi, kuten esimerkissä 25:

- (25) Tämä otetaan huomioon valmisteltavissa alusturvallisuutta koskevissa määräyksissä ja ohjeissa, joihin pyritään sisällyttämään määräykset turvallisuusarvioinnin (formal safety assessment) käyttämisen pakollisuudesta laivojen suunnittelussa, käytössä ja turvallisuusvalvonnassa. (hv98: 3)

Konkordanssihaku aineistosta *fi12* hakuehdolla (\* tuotti 81 tällaista esimerkkikontekstia. Hyvin usein (59 kertaa) suomenkielistä termiä seuraa sulkeissa englanninkielisestä termistä muodostettu lyhennetermi:

- (26) EGNOS on tärkeä tekninen osatekijä, jotta automaattinen tunnistusjärjestelmä (AIS) toimisi luotettavasti. (kl98: 70)

Englanninkielisen lyhennetermin lisäksi sulkeissa voi olla myös termin pitkä muoto, kuten esimerkissä 27:

- (27) Lähitulevaisuudessa navigoinnin perustietona on elektroninen kartta (ECDIS, Electronic Chart Display and Information System), - - (kiv97: 57)

Hakuehto (\* tuotti aineistosta *fi12* varsin runsaasti valideja osumia, huomattavasti enemmän kuin muut koetinedokkaat. Produktiivisuuden ansiosta sulkeet on kelvollinen synonymian koetin, mutta ongelma on roskaosumien runsaus. Tutkimukseni aineistossa sulkeet toimivat huomattavasti paremmin englanninkielisten vastineiden kuin suomenkielisten synonyymien koettimena. Validien osumien joukossa on yhteensä 140

sellaista osumaa, joissa ennen sulkeita on suomenkielinen termi ja sulkeissa englanninkielinen termi joko avattuna tai lyhenteenä. Sulkeet ovat siis vahva koetin, jos tavoitteena on hakea tekstistä englanninkielisiä vastineita tunnetuille suomenkielisille termeille. Koetin toimii myös toiseen suuntaan: sen avulla on mahdollista hakea suomenkielistä termiä tunnetulle englanninkieliselle termille. Englanninkielisten termien ripottelu suomenkieliseen tekstiin näyttää olevan merenkulun suomelle tyypillistä. Mahdollinen selitys on, että uudet merenkulun käsitteet tulevat suomeen englannin kielestä ja niitä nimeävät englanninkieliset termit ovat lukijoille tuttuja. Sen sijaan suomenkielinen termi saattaa olla vielä vakiintumaton. Laittamalla vakiintunut englanninkielinen termi sulkeisiin suomenkielisen termin jälkeen kirjoittaja selventää, mistä käsitteestä on kysymys.

Englannin kielen näkyvyyttä aineistossa kasvattavat edelleen ne 31 esimerkin 28 kaltaista tapausta, joissa sulkeissa on englanninkielinen termi ja välittömästi ennen sulkeita englanninkielinen lyhenne, ja ne 25 tapausta, joissa järjestys on päinvastainen, kuten esimerkissä 29:

- (28) Esimerkiksi IMO:n suosittama FMEA (Failure Mode and Effect Analysis) menetelmä on todettu käyttökelpoiseksi kun analysoidaan nopeiden alusten ohjailua ja kontrollijärjestelmiä. (smr02: 10)
- (29) Englanninkielissä maissa vastaava termi on Standard Operational Procedures (SOP). (kl98: 72)

Sulkeet ovat produktiivinen synonymian koetin, mutta vain murto-osa osumista on valideja. Poiminnan helpottamiseksi osumien kokonaismäärää on vähennettävä hakua tarkentamalla. Kokeilin ensin konkordanssihakua aineistosta *fi12* hakuehdolla (\*). Tämä hakuehto tuotti 200 osumaa. Osumalistaan tulostuneissa tapauksissa on sulkeissa usein numeroista ja kirjaimista koostuva merkkijono, esimerkiksi asiakirjan koodi:

- (30) IMO:n vaatimukset (Res.A.424) hyrräkompassista. (kiv97: 48)

Kun osumalistalta poistetaan tällaiset osumat, jäljelle jää 143 osumaa. Niiden joukossa on vielä esimerkiksi organisaatioista käytettyjä lyhenteitä mutta huomattavan paljon myös valideja osumia, joissa suluissa on käsitteestä käytettävä englanninkielinen lyhenne-termi, kuten esimerkissä 31, tai englanninkielinen termi, kuten esimerkissä 32:

- (31) Komentosiltayhteistyökoulutus (BRM) on käynnistynyt monessa varustamossa. (kiv97: 97)
- (32) Väsymys (fatigue) on tekijä, joka usein mainitaan merionnettomuuksien yhteydessä. (kiv97: 77)

Laajemman aineiston tekstien kirjoittajat käyttävät näissä tapauksissa sulkeita hyvin samalla tavalla kuin tekstin *pw02* kirjoittaja.

Hakuehdon (\*) tuottaman osumalistan tarkempi tarkastelu osoittaa, että noin joka toisessa eli 108 osumassa sulkeissa on isokirjainlyhenne, kuten edellä esimerkissä 31, tai merkkijono, jonka yhtenä osana on esimerkin tavoin isokirjainlyhenne. Kuitenkin näistäkin osumista noin neljäsosa on roskaa, sillä mukana on tapauksia, joissa isokirjainlyhenne ei ole termi vaan asiakirjan tai organisaation nimestä käytetty lyhenne, kuten esimerkissä 34:

- (33) Lisäksi ministerikokous päätti eräistä muista turvallisuutta lisäävistä toimenpiteistä joihin kuuluivat mm. satamatarkastusten valvonnan harmonisoiminen, erityisten herkkien alueiden selvittäminen (PSSA-alueet), turvallisuutta edistävän tiedon levittäminen, onnettomuuksien järjestelmällinen selvittäminen, riittävän hätähinauskapasiteetin ja öljyntorjuntakapasiteetin varmistaminen, hätäsatamien suunnittelu merihädässä oleville aluksille ja öljyntorjuntayhteistyön lisääminen erityisesti rantojen puhdistuksessa. (smr02: 92)
- (34) Suomen ympäristökeskus (SYKE) on tutkinut ympäristövahinkojen torjuntavalmiutta Suomessa (smr02: 29)

Olen luokitellut rosakaksi osumat, joissa sulkeissa on asiakirjan tai organisaation nimen lyhenne. Nämä tapaukset eivät kuitenkaan ole välttämättä täydellistä roskaa, sillä organisaatioiden ja asiakirjojen nimiä otetaan usein mukaan terminologisiin sanakirjoihin. Organisaatioiden ja asiakirjojen nimillä on terminologiassa ilmeisesti sijansa, sillä terminpoimintakokeeseen osallistuneet terminologit poimivat termilistoihinsa näitä selkeitä yksilökäsitteiden nimityksiä, vaikka määritelmän mukaan termi on yleiskäsitteen nimitys.

Isokirjainlyhenteen sisältävistä 108 osumasta 78 osumassa on sulkeissa käsitteen nimityksenä käytettävä isokirjainlyhenne, joka on yleensä englanninkielinen. Lyhenteen avattu muoto tai suomenkielinen termivastine on kaikissa näissä tapauksissa välittömästi ennen sulkeita:

- (35) Alusten automaattisen tunnistusjärjestelmän (AIS) käyttöönottoa halutaan edistää. (smr02: 92)

Näin sulkeiden avulla pystytään parhaassa tapauksessa poimimaan tekstistä englanninkieliset lyhenteet, niiden avatut muodot ja suomenkieliset vastineet.

Konkordanssihaku aineistosta *fi12* hakuehdolla (\* \*) tuotti 283 osumaa, joista 36 tapauksessa sulkeissa on kaksisanainen termi, lähes poikkeuksetta englanninkielinen. Suurimmassa osassa osumista sulkeissa on lähde, johon tekstin kirjoittaja viittaa. Huomattavan usein sulkeissa on joko sana *taulukko* tai sana *kuva*:

- (36) Kaikki tutkan näyttölaitteiden aiheuttamat ongelmat voitaisiin ratkaista siten, että tutka näyttäisi kompassilla vakautetun todellisen liikkeen "suunta ylös" (kuva 16). (kiv97: 54)

Konkordanssihaut hakuehdoilla (\* \* \*) ja (\* \* \* \*) tuottivat samansuuntaisia tuloksia. Hakuehto (\* \* \*), jonka täyttävissä osumissa suluissa on kolme sanaa, tuotti 140 osumaa, ja hakuehto (\* \* \* \*), jonka mukaisissa osumissa sulkeissa on neljä sanaa, tuotti 122 osumaa. Kummankin hakuehdon mukaisista osumista useimmissa sulkeiden sisällä on lähdetietoja, esimerkiksi (*Lazet ym. 1973*) tai (*Degani ja Wiener 1994*). Hakuehdot (\* \*), (\* \* \*) ja (\* \* \* \*) tuottivat kuitenkin myös osumia, joissa sulkeissa on englanninkielinen termi tai muuta terminologisesti kiinnostavaa tietoa. Hakuehdon rajaamisella niin, että sulkeiden sisällä on eripituisia lausumia, ei kuitenkaan saavuteta suurta hyötyä, sillä rajaukset eivät vähennä manuaalista työtä. Ainoa mielekäs rajausta on tehdä konkordanssihaku niin, että sulkeissa on yksi sana, jolloin osuimalistasta voidaan poimia isokirjainlyhenteet. Jos isokirjainlyhenteet tunnetaan tai ne on poimittu tekstistä esimerkiksi terminpoimintaohjelmalla, voi konkordanssihaun tehdä sijoittamalla sulkeisiin tunnetun lyhenteen, jolloin osuimalistaan tulostuvista konteksteista voidaan poimia termien avatut muodot. Jos osa termeistä on tunnettuja, voi konkordanssihaussa sulkeisiin sijoittaa tunnetun englanninkielisen termin, jolloin tekstistä voidaan poimia suomenkielisiä vastineita.



**Koetinehdokas eli.** Suomenkielisellä esitelmätekstillä *pw02* tehdyssä koetinpoiminnassa hyväksi synonymian koettimiksi osoittautui sulkujen lisäksi konjunktio *eli*. Konkordanssihakua aineistosta *fi12* hakuehdolla *eli* tuotti 75 osumaa, joista 27 osumassa konjunktio toimii termisynonymian koettimena. Roskan osuudeksi jää noin kaksi kolmasosaa osumista. Valideille tapauksille on yhteistä, että niissä kaikissa koetinehdokkaan *eli* kummallakin puolella on termi ja sen ja koetinehdokkaan välissä on korkeintaan yksi sana (esimerkiksi *lyhyesti*).

- (37) Komentosiltayhteistyön *eli* Bridge Resource Managementin (BRM) käsite on tullut Suomenkin kauppamerenkulkuun liikenneilmailun kautta. (kiv97: 63)

Esimerkissä 37 koetinehdokas *eli* ilmaisee suomenkielisen termin *komentosiltayhteistyö* ja englanninkielisen termin *Bridge Resource Management* välistä vastinesuhdetta, sillä kumpikin termi nimeää samaa käsitettä, mutta eri kielissä. Koska koetinehdokkaaseen *eli* näyttää valideissa tapauksissa aina liittyvän kaksi termiä, saadaan yhden tunnetun termin ja koetinehdokkaan yhdistelmähaulla poimittua tekstistä synonyymisiä termejä tai saman käsitteen erikielisiä nimityksiä.

**Yhteenvedo.** Suomenkielisessä aineistossani mahdollisista synonymian koettimista produktiivisin oli sulkeet. Ne ilmaisevat termisynonymiaa useimmiten termin pitkän ja lyhyen muodon välillä, mutta usein sulkeissa annetaan suomenkielisen termin englanninkielinen vastine. Sulkeet toimivatkin suomessa ennemmin termin ja vastineen koettimena kuin samankielisten synonyymisten termien koettimena. Roskaosumien runsauden vuoksi sulkeet toimivat koettimena parhaiten yhdessä tunnetun termin kanssa. Sulkeiden lisäksi suomenkielisestä aineistosta nousi esiin vain yksi synonymian terminologinen koetin, konjunktio *eli*. Tämä koetin ei ole yhtä produktiivinen kuin sulkeet, mutta kaikissa valideissa tapauksissa sen kummallakin puolella välittömässä läheisyydessä on termi.

#### 7.4.3 Määritelmän suomalaiset koettimet

Koettimista on todettu olevan hyötyä määritelmien haussa. Pearsonin (1998: 206) mukaan moni niiden avulla haettu määritelmä toimii sellaisenaan määritelmänä. Teksteissä voi olla terminologisia sisältömääritelmiä, joita Pearson (2000: 97) kutsuu muodollisiksi määritelmiksi (full explanation, formal definition). Niissä käsite on määritelty yläkäsitteen ja ainakin yhden erottavan piirteen avulla. Pearsonin mukaan nämä eivät ole niinkään harvinaisia teksteissä, mutta silti paljon useammin määritelmät ovat vajaita eli epämuodollisia määritelmiä (informal definition), joissa käsitettä kuvataan yhden olennaispiirteen avulla ilman yläkäsitettä. Teksteissä voi olla myös kontekstimääritelmiä joissa on esimerkiksi jokin looginen virhe tai selitteeseen kuuluvaa tietoa. Inkeri Vehmas-Lehto (henkilökohtainen tiedonanto 11.2.2006) nimittää tällaisia määritelmiä raakamääritelmiksi, koska ne voivat toimia lähtökohtana terminologiselle määritelmälle. Kaikkein yleisin on kuitenkin tapaus, jossa kiinnostava käsitetieto on tekstissä hajallaan tai haudattuna.

Olen käsitellyt erilaisia määritelmiä teoreettisesti jaksossa 7.1, jossa totesin, että tässä tutkimuksessa kiinnostuksen kohteena on kaikenlainen käsitteisiin liittyvä tieto, vaikka se ei olisikaan tarpeen määritelmän muodostamiseksi. Tässä jaksossa keskityn kuitenkin koetinehdokkaisiin, jotka ilmaisevat terminologisia tai vajaita määritelmiä, ja jätän hajautetun määritelmätiedon hakemisen tutkimukseni ulkopuolelle, sillä se vaatisi tekstin perinpohjaista manuaalista käsittelyä.

Koettimien käsinpoiminta tekstistä *pw02* tuotti vain yhden selkeän määritelmän koettimen, substantiivin *määritelmä* (ks. myös Pasanen 2005a: 295). Aineistosta *fi12* käsin tehdyssä poiminnassa mahdollisia määritelmän koettimia kuitenkin löytyi useita. Niistä tarkin oli verbi *tarkoittaa*, joka on koettimena toimiessaan joko aktiivimuodossa *tarkoittaa* tai passiivimuodossa *tarkoitetaan*. Verbin *tarkoittaa* lisäksi käsinpoiminta aineistosta *fi12* tuotti kaksi muuta leksikaalista määritelmän koetinehdokasta, verbin *ymmärtää* passiivimuodon *ymmärretään* ja konjunktion *eli*, sekä yhden mahdollisen välimerkki-koettimen, sulkeet. Konkordanssihaun tulokset olen koonnut taulukkoon 41.

TAULUKKO 41. Määritelmän terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
määritelmä	<i>määritelmä*</i>	16	14	88	vahva
määritellä	<i>määritellä*</i>	14	6	43	hyvä
tarkoittaa	<i>tarkoittaa/tarkoitetaan</i>	28	21	75	vahva
ymmärtää	<i>ymmärretään</i>	6	4	67	hyvä
eli	<i>eli</i>	75	25)	33	heikko
sulkeet	<i>(*</i>	821	9	1	heikko

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *määritelmä*.** Sana *määritelmä* esiintyy tekstissä *pw02* eri taivutusmuodoissa 10 kertaa. Kahdessa kontekstissa se toimii selkeästi määritelmän koettimena, kuten esimerkissä 38, ja kolmessa kirjoittaja selittää tietyn käsitteen määritelmää. Kuvaavaa on, että käsite on kaikissa viidessä tapauksessa *VTMIS* tai *VTS*:

- (38) Tähän kysymykseen on vaikea vastata lyhyesti, vaikka ensimmäinen lähellä oleva määritelmä saattaisi olla "VTMIS parantaa merikuljetusten tiedonkulkua". (*pw02*: 6)

Esimerkissä 38 kirjoittajan käsitteelle *VTMIS* antama määritelmä on vajaa, koska siitä puuttuu yläkäsitteen nimitys. Kirjoittaja on kuitenkin selvästi tarkoittanut lausuman määritelmäksi, koska se on laitettu lainausmerkkeihin.

Tekstin *pw02* kirjoittaja Paavo Wihuri käyttää esitelmässään sanaa *määritelmä* muissakin tehtävissä kuin määritelmän koettimena. Sanan yhteydessä ei aina ole käsitteen määritelmää, vaan sanaa käytetään esimerkiksi silloin, kun viitataan määritelmän muotoiluun tai määrittelyprosessiin tai kun määritelmä on annettu tekstissä aiemmin:

- (39) Vaikka asiasta keskustellaan vielä paljon, yhteistoimintakomitea on jäädyttänyt määritelmän, koska sen uskotaan olevan tarpeeksi kattava. (*pw02*: 6)

Tämän koettimen etu on, että se tuottaa vähän roskaa, sillä sen lähikontekstissa annetaan käsitteestä tietoa, vaikka ei kertaakaan muodollista terminologista määritelmää.

Aineistosta *fi12* määritelmien käsinpoiminta tuotti kahdeksan tapausta, joissa käsite on määritelty ja määritelmää ilmaisee koetinehdokas *määritellään*, *määritelmän mukaan*, *määritelmä*, voidaan *määritellä*. Kirjoittajat näyttäisivät siis käyttävän varsin harvoin tätä koetinta käsitteen määrittelyn yhteydessä. Koska koetinehdokas kuitenkin manuaalisen poiminnan perusteella tuottaa vähän roskaa, päätin testata sen toimivuutta konkordanssihaulla hakuehdolla *määritellä\*/määritelmä\**. Hakuehto tuotti 30 osumaa,

joista 20:ssa tekstin kirjoittaja antaa käsitteelle määritelmän, selittää sitä tai viittaa muualla tekstissä tai tekstin ulkopuolella annettuun määritelmään. Neljässä tapauksessa koetinehdokas toimii koettimena terminologiselle määritelmälle, jossa on yläkäsitteen nimitys ja erottava piirre:

- (40) Lain määritelmän mukaan luotsauksella tarkoitetaan alusten ohjailuun liittyvää toimintaa, jossa luotsi toimii aluksen päällikön neuvonantajana sekä vesialueen ja merenkulun asi-  
antuntijana. (smr02: 60)

Esimerkissä 40 määriteltävän käsitteen nimitys *luotsaus* on yhden sanan päässä koetinehdokkaan jälkeen. Määriteltävää käsitettä nimeävä termi on koetinehdokkaan lähikon-  
tekstissa kaikissa niissä tapauksissa, joissa käsitteelle annetaan määritelmä. Koetineh-  
dokkaan ja määriteltävän käsitteen nimityksen välissä on korkeintaan kaksi sanaa.

Useimmiten eli 14 validissa osumassa koettimena on substantiivi *määritelmä*. Sen  
sijaan mahdollinen verbikoetin *määritellä* tuotti yhdistetyn konkordanssihaun valideista  
osumista vain noin kolmanneksen, mutta kelvottomista osumista suurimman osan eli  
kahdeksan osumaa kymmenestä. Kelvottomien osumien määrää kasvattaa *määritellä*-  
verbin monimerkityksisyys. Kirjoittajat käyttävät verbiä *määritellä*, vaikka täsmälli-  
sempi olisi esimerkiksi verbi *määrittää*, *määrätä*, *tunnistaa*, *kuvata*, *päättää*, *kertoa* tai  
kuten esimerkissä 41 verbi *sopia*:

- (41) Toinen lainsäädäntöön liittyvä haaste on määritellä perustuslain edellyttämällä tavalla  
yksilöiden VTS-viranomaisen toimivalta ohjata alusliikennettä pysyvillä määräyksillä, - -  
(mv7: 22)

Kymmenessä tapauksessa kontekstissa selitetään määritelmää, joka on mahdollisesti  
annettu aikaisemmin tekstissä, tai viitataan tekstinulkoiseen määritelmään:

- (42) Kansainvälisen merenkulkujärjestön IMO:n määritelmä VTS:stä tekee selväksi, että asi-  
anomainen viranomais on mukana, ja että pääpaino on meriliikenteen turvallisuudella.  
(pw02: 4)

Koetinehdokas *määritellä* tai *määritelmä* ei tämän tutkimuksen aineiston perusteella ole  
kirjoittajakohmainen, vaikka kahdeksan tapausta 20:sta onkin yhdessä noin 2 000 sanan  
esitelmässä. Tämä selittyy sillä, että esitelmän aihe on käsitteiden *VTS* ja *VTMIS* määrit-  
tely. Muut tapaukset jakautuvat viiteen eri asiakirjaan, joten koetinehdokas *määritellä*  
tai *määritelmä* toimii koettimena joka toisessa aineistoon kuuluvassa asiakirjassa.

**Koetinehdokas tarkoittaa.** Konkordanssihaku aineistosta *f12* hakuehdolla *tarkoit-  
taa/tarkoitetaan* tuotti 28 osumaa, joista 21 on valideja. Kaikissa valideissa osumissa  
määriteltävää käsitettä nimeävä termi on ennen koetinehdokasta ja määritelmätieto seu-  
raa koetinehdokasta ja termiä. 13 osumassa termi on välittömästi ennen koetinehdokas-  
ta, kuten esimerkissä 43 termi *vakavuuslaskin*, seitsemässä osumassa termin ja koetin-  
ehdokkaan välissä on yksi sana, kuten esimerkissä 44 sana *mikä*:

- (43) Vakavuuslaskimella tarkoitetaan tietokoneohjelmaa, jonka avulla voidaan laskea laivan  
hydrostaattinen vakavuus. (smt99: 109)

- (44) Ne toimivat lähes poikkeuksetta "follow up"-periaatteella, mikä tarkoittaa sitä, että perä-  
sin asettuu samaan kulmaan kuin sen hallintavipu. (kiv97: 49)

Koetinehdokkaan *tarkoitetaan* tuottamat tapaukset jakautuvat viiteen asiakirjaan, siten että suurin osa tapauksista on kahdesta laajimmasta asiakirjasta. Tästä voi päätellä, että koetinehdokas ei ole kirjoittajakohtainen.

**Koetinehdokas *ymmärtää*.** Verbin *ymmärtää* passiivimuoto *ymmärretään* on manuaalisen koetinhaun perusteella mahdollinen määritelmän koetin. Konkordanssihakua hakuehdolla *ymmärretään* tuotti kuusi osumaa, joista neljässä koetinehdokas toimii määritelmän koettimena:

- (45) Vakavuudella ymmärretään kunnossa olevan laivan kykyä säilyttää pystyasentonsa tilapäisistä tai toistuvista poikkeamista huolimatta. (smt99: 16)

Esimerkissä 45 määriteltävä käsite on ominaisuuskäsite *vakavuus*, jonka termi on välittömästi ennen koetinehdokasta ja määritelmä seuraa välittömästi koetinehdokkaan jälkeen. Määriteltävän käsitteen nimitys onkin kaikissa valideissa tapauksissa yhtä lukuun ottamatta välittömästi ennen koetinehdokasta. Koetinehdokas *ymmärretään* tuottaa keltottomia osumia samasta syystä kuin verbi *määritellä*: myös *ymmärretään* on monimerkityksinen.

**Koetinehdokas *eli*.** Toisella käsinpoiminnan tuottamalla määritelmän koetinehdokkaalla konjunktioilla *eli* tehty konkordanssihakua aineistosta *fi12* tuotti 75 osumaa. Määritelmän koettimena tämä koetinehdokas näyttäisi tuottavan runsaasti roskaa, sillä osumista vain yhdessä kolmasosassa eli 25 osumassa annetaan määritelmätietoa jostakin merenkulun turvallisuuden käsitteestä:

- (46) Tässä esitetyt tulokset kuvaavat IMDG-paloskenaarioihin liittyvää jäännösriskiä eli riskiä, joka jää käytännössä jäljelle vielä näiden toimenpiteiden toteuttamisen jälkeen. (smt99: 93–94)

Esimerkissä 46 määriteltävän käsitteen termi *jäännösriski* on välittömästi ennen koetinehdokasta, jota määritelmä seuraa. Järjestys on sama 18 muussakin tapauksessa. Neljässä tapauksessa määriteltävän käsitteen nimitys on koetinehdokkaan jälkeen:

- (47) Päällikön oli käytettävä paikalliset vedet tuntevaa henkilöä eli luotsia apunaan siellä, missä luotsipalvelua oli järjestetty. (kiv97: 68)

Esimerkissä 47 määritelmä *paikalliset vedet tunteva henkilö* on ennen koetinehdokasta ja määriteltävän käsitteen nimitys *luotsi* koetinehdokkaan jälkeen. Kelpaamattomissa osumissa *eli*-konjunktioita on saatettu käyttää esimerkiksi sekä numeroina että prosentteina ilmaistun numeerisen arvon yhteydessä, kuten esimerkissä 48:

- (48) Suomessa vuonna 1999 käyneistä 209 säiliöaluksista kuului korkeimpaan jääluokkaan, 1A Super, yhteensä 9 kpl, eli 4.3 % (MKL, 1999). (smr02: 26)

**Koetinehdokas *sulkeet*.** Leksikaalisten koetinehdokkaiden lisäksi manuaalinen poiminta aineistosta *fi12* tuotti yhden mahdollisen välimerkkikoettimen, sulkeet. Sulkeet on niin sanottu yleiskoetin, joka saattaa ilmaista määritelmän lisäksi ainakin synonymiaa, lähikäsitteitä, geneeristä käsitesuhdetta tai koostumussuhdetta. Vaikka sulkeet eivät ole käsinpoiminnan perusteella ensisijaisesti määritelmän koetin, tuotti käsinpoiminta kuitenkin niin paljon tapauksia, että sulkeiden toimintaa määritelmän koettimena kannattaa tutkia. Konkordanssihakua aineistosta *fi12* hakuehdolla (\*) tuotti 821 osumaa, joista 30

tapauksessa koetinehdokas toimii määritelmän koettimena. Sulkeissa saattaa olla termi (joskus englanninkielinen), määritelmä tai selite. Jos sulkeissa on termi, määritelmä tai selite edeltää termiä:

- (49) Toiseen luokkaan luetaan vastaavansuuruiset säiliöalukset, jotka on varustettu suojaavin ja osastoiduin painolastitankein (MARPOL-tankkerit). (smr02: 88)

Useimmiten eli 19 tapauksessa 30:sta määritelmän ja termin järjestys on sama kuin esimerkissä 49, jossa määritelmä on ennen sulkeissa olevaa termiä. Sulkeissa olevista 19 termistä 10 on englanninkielisiä. Jos taas sulkeissa on määritelmä, termi on annettu sitä ennen. ”Ylimääräisiä” sanoja ei välissä ole:

- (50) Tämä johtuu siitä, että Perämerellä yhtenäisen jään (tasainen jää ja vallien yhteen jäätynyt osa) paksuus on suurempi suuremmasta pakkasummasta johtuen. (smr02: 37)

Termi on ennen sulkeita ja määritelmä sulkeissa noin kolmanneksessa tapauksia eli 11 tapauksessa 30:sta. Määritelmiä voi aineistosta hakea laittamalla tunnetun termin sulkeisiin, jolloin käsitteen määritelmä saattaa olla konkordanssihaun tuottamassa kontekstissa ennen sulkeita. Toinen mahdollisuus on hakea määritelmää tunnetun termin ja vasemman sulkeen yhdistelmällä, jolloin mahdollisesti sulkeissa oleva määritelmä tulostuu konkordanssihaun osumalistaan. Ainakaan määritelmän koetinehdokkaana sulkeet eivät ole kirjoittajakohmainen välimerkki, sillä validit osumat ovat seitsemästä eri asiakirjasta.

**Koetinehdokas olla.** Pearsonin (1998: 104, 140) mukaan englannin kielessä verbi *be* (*is/are*) on leimallisesti sisältömääritelmän koetin ja jopa ylivoimaisesti yleisin sisältömääritelmän koettimista. Jos tämä englanninkielisellä aineistolla saatu tutkimustulos on yleistettävissä suomeen, pitäisi *olla*-verbin muotojen *on* ja *ovat* olla hyviä sisältömääritelmän koettimia. Tutkimukseni suomalaisessa aineistossa on käsinpoiminnan perusteella ainakin yhdeksän tapausta, joissa *olla*-verbin muotoa *on* seuraa formaalinen sisältömääritelmä, jossa määritelmä on muotoa X on Y + relatiivipronominilla alkava sivulause. Kuudessa tapauksessa määriteltävää käsitettä nimeävä termi on välittömästi ennen koetinehdokasta:

- (51) Merenkulkulaitos on meripelastusviranomainen, jonka tehtävänä on valvoa merialuetta alusliikennepalvelujärjestelmän avulla onnettomuus- ja vaaratilanteiden havaitsemiseksi sekä osallistua etsintä- ja pelastustoimintaan. (mv7: 23)

Muotovaatimukset täyttävää sisältömääritelmää tavallisempi tapaus aineistossa on kuitenkin vajaa määritelmä, jossa on jokin käsitepiirre, mutta lähin yläkäsite puuttuu:

- (52) Laivaturvallisuus on kuitenkin olemukseltaan dynaaminen eikä staattinen, - - (lat97: 2)

Käsinpoiminta aineistosta *fi12* tuotti 20 vajaata määritelmää, joiden lähikontekstissa on koetinehdokas *on* tai *ovat*. Kaikissa tapauksissa yhtä lukuun ottamatta määriteltävää käsitettä nimeävä termi on välittömästi ennen koetinehdokasta *on* tai *ovat*. Käsinpoiminnan perusteella koetinehdokas voi olla myös perusmuodossa *olla*, jolloin sitä edeltää modaaliapuverbi (voi, tulee). Konkordanssihaku tällä koetinehdokkaalla ilman tunnettua termiä ei ole mielekäästä, sillä *olla*-verbi tuottaa runsaasti roskaa: pelkästään hakuehto *on* tuottaa aineistosta 2 053 osumaa. Koska *olla*-verbi näyttää määritelmän koettimena olevan lähes aina heti termin perässä, se kannattaa kuitenkin hyväksyä mää-

ritelmän suomalaisten koettimien joukkoon, koska sillä saadaan runsaasti määritelmätietoa, josta osa on formaalisina sisältömääritelminä.

**Yhteenveto.** Substantiivi *määritelmä* sekä verbi *tarkoittaa* osoittautuivat tutkimukseni suomalaisessa aineistossa vahvoiksi määritelmän koettimiksi, sillä ne tuottivat vähän roskaa eivätkä näyttäisi olevan kirjoittajakohtaisia. Verbi *määritellä* tuotti lähes yhtä monta osumaa kuin substantiivi *määritelmä*, mutta sen tarkkuus on huomattavasti alhaisempi. Nämä koettimet tuottivat vajaita määritelmiä, sillä niissä mainitaan jokin käsitepiirre, mutta yläkäsite on mainittu 20 validista tapauksesta vain viidessä. Näistä neljässä yläkäsite on yleinen tieteen termi (prosessi, toiminta, kokonaisuus). Verbi *tarkoittaa* tuotti runsaasti valideja tapauksia ja vähän roskaa ja osoittautui siten erittäin vahvaksi määritelmän koettimeksi, vaikka vain kolmasosassa valideista tapauksista on mainittu määriteltävän käsitteen yläkäsite. Verbi *tarkoittaa* ei myöskään näytä olevan kirjoittajakohtainen.

#### 7.4.4 Geneerisen käsitesuhteen ylä- ja alakäsitteen suomalaiset koettimet

Geneerisellä käsitesuhteella tarkoitan käsitesuhdetta, jossa alakäsitteellä on kaikki yläkäsitteen käsitepiirteet ja lisäksi vähintään yksi erottava piirre. Geneeristä käsitesuhdetta pidetään tärkeimpänä käsitesuhdetyyppinä, joten ei ihme, että juuri geneerisen käsitesuhteen koettimet ovat aiemmissa tutkimuksissa saaneet eniten huomiota (ks. esim. Ahmad & Rogers 2001; Bowker & Pearson 2003; Grinstead 2000; Kavanagh 1995; Pearson 1996).

Suomenkielisellä esitelmätekstillä *pw02* tehty konkordanssihaku ja koettimien käsinpoiminta aineistosta *fi12* tuottivat runsaasti mahdollisia geneerisen suhteen koettimia. Niitä ovat komparatiivinen alistuskonjunktio *kuten*, substantiivit *esimerkki* ja *tyyppi*, jakonjunktio yhdessä sanojen *sekä*, *kuten*, *esimerkiksi*, *muuta*, *muiden*, *muut* kanssa, sanaliitto *kuten esimerkiksi*, lyhenne *mm.* sekä verbit *sisältyä*, *sisältää*, *jakaa*, *jaotella* ja *luokitella*. Mahdollisia geneerisen suhteen välimerkkikoettimia ovat käsinpoiminnan perusteella sulkeet ja kaksoispiste. Lisäksi käsinpoiminta tuotti runsaasti sellaisia mahdollisia geneerisen suhteen koettimia, jotka esiintyvät lähtötekstissä vain muutaman kerran. Tällaisia ovat substantiivit *luokka* ja *ryhmä* sekä verbit *laskea*, *lukea*, *kuulua* *johonkin*, *tarkoittaa* ja *muodostaa*. Koetinedokkaiden testaamiseksi tein konkordanssihaun koetinaineistosta *fi12* niillä koetinedokkailla, jotka ovat tunnetuilla termeillä tehdyn konkordanssihaun ja käsinpoiminnan perusteella todennäköisesti parhaita geneerisen suhteen koettimia. Konkordanssihakujen tulokset ovat taulukossa 42.



TAULUKKO 42. Geneerisen käsitesuhteen terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
kuten	<i>kuten</i>	76	34	45	hyvä
esimerkki	<i>kuten esimerkiksi</i>	8	5	63	hyvä
muu	<i>ja muut/ja muiden</i>	20	12	60	vahva
	<i>ja muita/myös muita</i>	9	4	44	heikko
tyyppi	<i>*tyyppi/*tyypit</i>	15	7	47	hyvä
mm.	<i>ovat mm.</i>	7	4	57	hyvä
sisältyä, sisältää	<i>sisältyy/sisältä*</i>	62	5	8	heikko
jakaa	<i>voidaan jakaa/jaettu</i>	20	6	30	heikko
luokitella	<i>luokitel*</i>	12	5	42	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *kuten*.** Aineistosta *pw02* tehdyn konkordanssihaun perusteella alistuskonjunktio *kuten* ottaa seurakseen sanan *esimerkiksi* geneerisen suhteen koettimena toimiessaan. Käsinpoiminta aineistosta *fi12* osoitti kuitenkin, että *kuten* voi toimia koettimena myös ilman seuralaista. Tästä syystä tein konkordanssihaun aineistosta *fi12* hakuehdolla *kuten*. Haku tuotti 76 osumaa, joista 34:ssä *kuten* toimii geneerisen suhteen koettimena. Roskaa osumista on siis hieman yli puolet. Kaikissa niissä tapauksissa, joissa *kuten* on geneerisen suhteen koetin, yläkäsitteen nimitys on tekstissä välittömästi ennen koetinehdokasta, kuten esimerkissä 53 *päästö mereen*:

- (53) Itse liikenteen keskeisimmät vaikutusmuodot ovat emissiopäästöt ja erilaiset päästöt mereen, kuten harmaa vesi, musta vesi, pilssivesi ja painolastivesi. (smr02: 55)

Geneerisen suhteen alakäsitteen nimitys tai nimitykset ovat nekin tekstissä yhtä esimerkkikontekstia lukuun ottamatta välittömästi koetinehdokkaan jälkeen, kuten esimerkissä 53 alakäsitteiden nimitykset *harmaa vesi*, *musta vesi*, *pilssivesi* ja *painolastivesi*. Joissakin tapauksissa yläkäsitteen nimitys ei ole merenkulun turvallisuustermi vaan luokkasana, kuten *toimija* esimerkissä 54:

- (54) Sen käyttäjinä ovat kaikki satamaympäristöön liittyvät toimijat, kuten merenkuluviranomaiset, tulli, alusten edustajat, ahtausliikkeet, satamat jne. (smr02: 68)

Esimerkissä 54 yläkäsitteellä *toimija* on sataman käsitejärjestelmässä lukuisia alakäsitteitä, joista esimerkikikontekstissa on mainittu viisi. Lyhenne *jne.* luettelon jälkeen osoittaa, että luettelo ei ole täydellinen, vaan alakäsitteitä voi olla muitakin. Koetinehdokas *kuten* toimii myös vieruskäsitteiden koettimena, sillä 23 tapauksessa koetinehdokasta seuraa luettelo, jossa on vähintään kaksi käsitteen nimitystä.

**Koetinehdokas *esimerkki*.** Substantiivista *esimerkki* manuaalisen poiminnan perusteella muotoiltu hakuehto *esimerkiksi* tuotti 165 osumaa. Määrä on moninkertainen verrattuna samasta aineistosta *fi12* tehtyyn käsinpoimintaan, sillä tämä tuotti vain neljä kontekstia, joissa sana *esimerkiksi* toimii geneerisen suhteen koettimena. Tästä voidaan päätellä, että sana *esimerkiksi* yksinään tuottaa aivan liian paljon roskaa, jotta se voitaisiin hyväksyä toimivaksi geneerisen suhteen koettimeksi. Kahdeksassa tapauksessa välittö-

mästi ennen koetinehdokasta on alistuskonjunktio *kuten*. Yhdistämällä koetinehdokkaat *kuten* ja *esimerkiksi* samaan hakuehtoon *kuten esimerkiksi* saadaan osumien määrä supistumaan kahdeksaan osumaan, joista viidessä ennen koetinehdokasta on yläkäsitteen nimitys ja välittömästi koetinehdokkaan jälkeen vähintään yhden alakäsitteen nimitys:

- (55) Huolimatta jäljempänä esitetyistä laajemmista viittauksista "Yhteisiin palveluihin", joita tavallisesti kuvataan suoranaisesti laivoihin liittyvinä palveluina kuten esimerkiksi luotsi ja hinauspalveluihin jne., ei ole selvästi ilmaistu, että vaadittaisiin merikuljetuksen tehokkuutta. (pw02: 4)

Esimerkissä 55 geneerisen käsitesuhteen yläkäsitteen nimitys *laivoihin liittyvä palvelu* on välittömästi ennen koetinehdokasta ja sen alakäsitteiden nimityksistä kaksi, *luotsi* ja *hinauspalvelu* välittömästi koetinehdokkaan jälkeen.

**Koetinehdokas *muu*.** Seuraavaksi tein konkordanssihaun hakuehdolla *muut/muita/muihin*, joka tuotti 55 osumaa. Hakuehto oli muotoiltu käsinpoiminnan perusteella, mutta konkordanssihaun tulosten runsaus osoitti, että hakua on tarkennettava. Konkordanssihaun tulosten perusteella tarkensin hakuehdon muotoon *ja muut/ja muiden*. Näin osumien määrä supistui 20 osumaan, joista 12 on valideja esimerkkejä koetinehdokkaan toiminnasta geneerisen suhteen koettimena:

- (56) Pelastuspalvelu ja muut korjaavat toimenpiteet (smt99: 14)

Koetinehdokas *ja muut* poikkeaa edellä tarkastelluista koetinehdokkaista *kuten* ja *esimerkiksi* siten, että yläkäsitteen nimitys on nyt tekstissä koetinehdokkaan jälkeen, kuten *korjaava toimenpide* esimerkissä 56, ja alakäsitteen nimitys ennen koetinehdokasta, kuten *pelastuspalvelu* samassa esimerkissä. Koetinehdokas *ja muut* voi myös johtaa harhaan, kuten esimerkissä 57:

- (57) Käsiteltävät onnettomuustyyppit ovat: yhteentörmäys, karilleajo, törmäys (muuhun esineeseen kuin laiva tai kari) ja muut vuototapaukset. (smt99: 109)

Jos sanaliitto *ja muut* olisi geneerisen suhteen koetin esimerkissä 57, niin yläkäsite olisi *vuototapaus* ja sen alakäsitteitä olisivat *yhteentörmäys*, *karilleajo* ja *törmäys (muuhun esineeseen kuin laiva tai kari)*. Näinhän ei kuitenkaan ole, vaan kaikille yhteinen yläkäsite on tekstissä aiemmin mainittu *onnettomuus*.

Alkuperäisen hakuehdon tuottamien tulosten perusteella korvasin hakuehdon *muut/muita/muihin* toisen osan, sanan *muuta* hakuehdolla *ja muita/myös muita*, jolloin osumien määrä väheni yhdeksään. Niistä neljä on validia esimerkkiä koetinehdokkaan toiminnasta geneerisen suhteen ala- ja yläkäsitteen koettimena. Kaikissa valideissa osumissa yläkäsitteen nimitys on tekstissä välittömästi koetinehdokkaan jälkeen:

- (58) Talouselämän voimistuessa ja kaupankäynnin lisääntyessä on tärkeää, ettei laivaliikennettä ja muita liikennejärjestelmiä rajoita mitkään esteet tai pullonkaulat. (smr02: 12)

Esimerkissä 58 yläkäsitteen nimitys *liikennejärjestelmä* on välittömästi koetinehdokkaan jälkeen, ja tässä esimerkissä myös alakäsitteen nimitys *laivaliikenne* on välittömästi ennen koetinehdokasta, mutta alakäsitteen nimitys voi olla myös konkordanssilistaan tulostuneen kontekstin ulkopuolella.

Alkuperäisen hakuehdon *muut/muita/muihin* kolmas osa, sana *muihin*, osoittautui konkordanssihaun osumien perusteella ”köyhäksi” koetinehdokkaaksi, sillä sen tuotta-

mien 11 osuman joukossa oli vain yksi validi tapaus geneerisestä suhteesta eikä sekään ole erityisen edustava:

- (59) Merkintälaskua verrataan koko ajan muihin paikanmäärityslaitteisiin. (kiv97: 52)

Esimerkissä 59 alakäsité on *merkintälasku*, jonka yläkäsité on kirjoittajan mukaan *paikanmäärityslaitte*. Yläkäsité on kuitenkin *paikanmääritysmenetelmä*. Tutkimukseni aineistossa koetinehdokkaan tarkkuus on korkein yhdistetyllä hakuehdolla *ja muut/ja muiden/ja muita/myös muita*, jonka tuottamista 29 osumasta 16 on valideja.

**Koetinehdokas *tyyppi*.** Substantiiveista *tyyppi* on käsinpoiminnan perusteella mahdollinen geneerisen suhteen koetin. Konkordanssihaku aineistosta *fi12* hakuehdolla *\*tyyppi/\*tyypit* tuotti 15 osumaa, joista seitsemän on valideja. Koetinehdokas ei kuitenkaan kasvata terminologisesti kiinnostavan tiedon määrää kovinkaan paljoa, sillä tällä hakuehdolla saadaan tutkimukseni aineistosta poimittua vain yhden yläkäsitteen ja sen alakäsitteiden väliset geneeriset suhteet:

- (60) Karilleajot ja pohjakosketukset ovat suomalaisalusten yleisin vaurioita aiheuttava onnettomuustyyppi. (smt99: 55)

Tämä yläkäsité on *onnettomuus*, ja alakäsitteet ovat *tulipalo*, *myrkyllisen aineen vuoto*, *yhteentörmäys*, *karilleajo*, *törmäys muuhun esineeseen kuin laiva tai kari*, *vuoto* ja *pohjakosketus*. Valideista esimerkeistä kuusi on samasta asiakirjasta, joten koetinehdokas näyttää olevan kirjoittajakohtainen. Muissa kuin valideissa osumissa koetinehdokas on joko yhdyssanan alkuosana, jolloin loppuosa on yleensä *koulutus* tai *hyväksyntä*, tai yhdyssanan jälkimmäisenä osana, jolloin alkuosa on joko *proto*, *alus* tai *laiva*:

- (61) Lähtökohdan laivan rakenneturvallisuudelle muodostaa asiakkaan, laivan omistajan asettamat toiminnalliset vaatimukset, joita ovat mm. laivatyyppi, lastit ja niiden käsittely sekä purjehdusalue. (lat97: 16)

Esimerkin 61 tavoin sana *laivatyyppi* tai synonyyminen sana *alustyyppi* viittaa yleensä alusten luokitteluun ja on pikemmin termin kuin koettimen roolissa. Substantiivi *tyyppi* on aineistossani yleinen sanan osana, sillä hakuehto *\*tyyp\** tuotti 84 osumaa, joissa yhdyssanan yhtenä osana on *tyyppi*. Se ei kuitenkaan ole hyvä geneerisen suhteen koetin, koska sen käyttö koettimena näyttäisi olevan kirjoittajakohtaista.

**Koetinehdokas *mm*.** Käsinpoiminta tuotti yhden mahdollisen geneerisen suhteen lyhennekoettimen, lyhenteen *mm*. Lyhenne osoittautui yleiseksi tutkimukseni aineistossa, sillä konkordanssihaku hakuehdolla *mm*. tuotti 72 osumaa. Osumien määrä on liian suuri manuaalisesti käsiteltäväksi, joten hakuehtoa on tarkennettava. Kun hakuehto muutetaan osumalistan perusteella muotoon *ovat mm.*, konkordanssilistaan tulostuu seitsemän osumaa, joista neljä on valideja geneerisen suhteen tapauksia:

- (62) Tällaisia liikennerajoituksia ovat mm. alusten jääluokat, joilla estetään huonokuntoisten tai jäissäkulkumuinaisuuksiltaan heikkojen alusten joutuminen vaikeuksiin. (smr02: 98)

Esimerkissä 62 alakäsitteen nimitys *aluksen jääluokka* on tekstissä välittömästi koetinehdokkaan *ovat mm.* jälkeen, kuten muissakin valideissa tapauksissa.

**Koetinehdokkaat *sisältyä* ja *sisältää*.** Manuaalinen koetinhaku tuotti viisi mahdollista geneerisen suhteen verbikoetinta: verbit *sisältyä*, *sisältää*, *jakaa*, *jaotella* ja *luokitella*. Manuaalisen haun tulosten perusteella tein konkordanssihaun ensin hakuehdolla *sisält\**, jotta hakuehto kattaisi kaikki relevantit *sisältyä* ja *sisältää* verbien taivutusmuodot. Konkordanssihaku aineistosta *fi12* tällä hakuehdolla tuotti 88 osumaa. Suurin osa osumista oli roskaa, joten hakuehtoa oli tarkennettava. Tarkennettu hakuehto *sisältyy/sisältää\** tuotti 62 osumaa, mutta konkordanssilistan lähempi tarkastelu osoitti, että listassa on hyvin vähän tapauksia, joissa käsitteiden välillä on geneerisen ylä- ja alakäsitteen suhde:

- (63) Radioliikenteen vakiorutiinit (Radio Procedures) sisältävät hätäliikenteen standardi-ilmoitukset matkustajille ja miehistölle sekä tiivistetyt VHF-, MF-, ja HF- radiopuhelimien sekä DSC (Digital Selective Call) -laitteiden toimintaohjeet. (kiv97: 62)

Esimerkissä 63 on geneerinen käsitesuhde, jonka yläkäsite on *radioliikenteen vakiorutiini* ja alakäsitteet *hätäliikenteen standardi-ilmoitus matkustajille*, *hätäliikenteen standardi-ilmoitus miehistölle*, *VHF-radiopuhelimien toimintaohjeet*, *MF-radiopuhelimien toimintaohjeet*, *HF-radiopuhelimien toimintaohjeet* ja *DSC-laitteiden toimintaohjeet*. Esimerkissä on siis kuusi alakäsitettä, sillä matkustajille ja miehistölle annetaan erilainen standardi-ilmoitus ja erilaisilla radiolaitteilla on omat toimintaohjeet. Tässäkin esimerkissä kirjoittaja on käyttänyt alakäsitteistä elliptisiä ilmauksia.

Geneerisen ylä- ja alakäsitesuhteen tapausten sijasta hakuehdolla *sisält\** tehty konkordanssihaku tuotti konteksteja, joissa on koostumussuhteen ylä- ja alakäsitteen välinen suhde (ks. jakso 7.4.5). Tulos on yllättävä, sillä englanninkielisillä aineistoilla tehtyjen tutkimusten perusteella verbi *include*, joka on suomenkielisen verbin *sisältää* vastine, on pikemmin geneerisen suhteen kuin koostumussuhteen koetin (ks. Bowker & Pearson 2002: 219; Grinstead 2000: 40–42; Kavanagh 1995). Suomen verbit *sisältyä* ja *sisältää* ovat kuitenkin usein juuri koostumussuhteen koettimia. Esimerkiksi kun kirjoittaja sanoo, että rintaontelo sisältää sydämen ja keuhkot, hän on täysin oikeassa, sillä sydän ja keuhkot ovat rintaontelon sisällä, mutta sydän tai keuhkot eivät ole rintaontelon geneerisiä alakäsitteitä. Verbin *sisältää* käyttökelpoisuus koettimena kärsii siitä, että se voi olla joko koostumussuhteen tai geneerisen suhteen koetin ja kirjoittajat käyttävät sitä huolimattomasti.

**Koetinehdokas *jakaa*.** *Jakaa*-verbin hakuehdon rajasin käsinpoiminnan perusteella muotoon *voidaan jakaa/jaettu*. Hakuehto tuotti 20 osumaa, joista 18:ssa verbi ilmaisee joko geneerisen suhteen tai koostumussuhteen ylä- ja alakäsitteen välistä suhdetta. Valideista osumista suurimmassa osassa, 12 osumassa, käsitteiden välillä on koostumussuhde (ks. jakso 7.4.5). Kuudessa validissa osumassa on geneerinen suhde:

- (64) Vaaralliset aineet voidaan jakaa irtolasteina IBC-koodin määräysten mukaisesti kuljetettaviin aineisiin ja pakattuihin eli IMDG-koodin määräysten mukaisesti kuljetettaviin aineisiin. (smt99: 76)

Esimerkissä 64 geneerisen suhteen yläkäsitteen nimitys *vaarallinen aine* on tekstissä välittömästi ennen koetinehdokasta ja alakäsitteiden nimitykset *IBC-koodin määräysten mukaisesti kuljetettava aine* ja *IMDG-koodin määräysten mukaisesti kuljetettava aine* välittömästi koetinehdokkaan jälkeen. Muissa geneerisen suhteen tapauksissa yläkäsitteen nimitys on yleensä välittömästi ennen koetinehdokasta mutta alakäsitteiden nimitykset saattavat jäädä konkordanssilistan tulostusalueen ulkopuolelle, kuten esimerkissä 65:

- (65) Seuraavassa toimintaan liittyvät turvallisuus- tai vaaratekijät on jaettu neljään pääryhmään, jotka koostuvat joukosta erityisiä tapahtumatyyppejä. (smt99: 16)

Esimerkissä 65 yläkäsitteiden nimitykset *turvallisuustekijä* ja *vaaratekijä* ovat yhden sanan päässä ennen koetinehdokasta *jaettu*, mutta alakäsitteiden nimitykset tulevat tekstissä myöhemmin, neljästä alakäsitteestä viimeisinä käsiteltävien nimitykset vasta monen sivun jälkeen.

**Koetinehdokas *luokitella*.** *Sisältää*-, *sisältyä*- ja *jakaa*-verbit ovat yleisluontoisia ylä- ja alakäsitteen välisen suhteen koettimia, sillä suppean kontekstin perusteella ei aina pysty sanomaan, onko käsitteiden välillä geneerinen suhde vai koostumussuhde. Verbit *sisältää* ja *sisältyä* ovat konkordanssilistan validien osumien perusteella ensi sijassa koostumussuhteen koettimia. Verbi *jakaa* sen sijaan näyttäisi toimivan sekä geneerisen että koostumussuhteen koettimena. Verbi *luokitella* puolestaan on selvästi geneerisen suhteen koetin. Hakuehto *luokitel\** tuotti 12 osumaa, joista viisi on valideja geneerisen suhteen tapauksia:

- (66) Pakatut vaaralliset aineet luokitellaan yhdeksään pääryhmään, jotka on esitetty taulukossa 4. (smt99: 77)

Esimerkissä 66 yläkäsitteen nimitys *pakattu vaarallinen aine* on tekstissä välittömästi ennen koetinehdokasta *luokitellaan*, mutta alakäsitteiden nimitykset on haettava tekstiin liittyvästä taulukosta. Seitsemässä osumassa ei ole ylä- ja alakäsitteiden välistä käsitteesuhdetta, vaan niissä puhutaan aineiden, lastien, yksiköiden tai riskien luokittelusta yleensä:

- (67) Vaaralliseksi luokiteltujen aineiden lukumäärä on suuri ja se kasvaa ilmeisesti jatkuvasti. (smt99: 82)

**Koetinehdokas *kaksoispiste*.** Koettimien käsinpoiminnan perusteella sulkeet ja kaksoispiste ovat mahdollisia geneerisen suhteen koettimia. Konkordanssihaku aineistosta *fi12* hakuehdolla \*: tuotti peräti 225 osumaa. Joukossa on ilmeisesti runsaasti kelpaamattomia osumia, sillä käsinpoiminnassa samasta aineistosta löytyi vain kahdeksan kontekstia, joissa kaksoispiste toimii geneerisen suhteen koettimena. Valideja osumia voi poimia osumalistasta tunnettujen termien avulla, sillä valideista osumista viidessä yläkäsitettä nimeävä termi on välittömästi ennen kaksoispistettä ja alakäsitettä nimeävä termi tai muu käsitettä nimeävä ilmaus kaikissa tapauksissa välittömästi kaksoispisteen jälkeen:

- (68) Tarkastelu tehdään kolmelle eri liikennealueelle: kansainvälisille vesille, poikkeaminen kohti määräsatamaa ja kansallisille vesille. (smr02: 42)

Esimerkissä 68 yläkäsitteen nimitys *liikennealue* on välittömästi ennen kaksoispistettä ja alakäsitteiden nimitykset välittömästi kaksoispisteen jälkeen. Alakäsitteiden nimityksistä *kansainväliset vedet* ja *kansalliset vedet* ovat puhekielisiä nimityksiä ja *poikkeaminen kohti määräsatamaa* on ilmeisen pätevä ilmaus kontekstissaan, vaikka se muissa konteksteissa tuskin on *liikennealueen* alakäsite eli eräänlainen liikennealue.

**Koetinehdokas *sulkeet*.** Toinen mahdollinen geneerisen käsitesuhteen välimerkki-koetin, sulkeet, tuotti hakuehdolla (\* 25 sellaista osumaa, joissa on yläkäsitteen ja aina-

kin yhden alakäsitteen välinen geneerinen käsitesuhde. Kaikissa konkordanssihaun tuot-  
tamissa valideissa osumissa yhtä lukuun ottamatta yläkäsitteen nimitys on ennen sulkei-  
ta ja alakäsitteen tai alakäsitteiden nimitykset ovat sulkeissa. 25 osumasta 21:ssä yläkä-  
sitteen nimitys on välittömästi ennen sulkeita, kuten esimerkissä 69 yläkäsitteen nimitys  
*säiliöalus*:

- (69) Tarkempaa selvitystä vaatisivat mm. eri jääluokkiin kuuluvien alusten liikennöinti Suo-  
menlahdella, kaksoispohjan- ja rungon yleisyyden sekä erityyppisten ja -kokoisten säiliö-  
alusten (raakaöljy-, tuote- ja kemikaali- sekä kaasutankkereiden) osuukien selvitys koko  
Suomenlahden tankkeriliikenteessä. (smr02: 27)

**Koetinehdokas *olla*.** Englanninkielisillä aineistoilla tehtyjen tutkimusten perusteella  
*be*-verbin preesensmuodot *is* ja *are* yhdessä artikkelien *a*, *an* tai *the* kanssa ovat hyviä  
geneerisen suhteen koettimia (ks. Ahmad & Rogers 2001: 749; Bowker & Pearson  
2002: 219; Feliu & Cabré 2002: 5; Grinstead 2000: 40–42; Kavanagh 1995; Pearson  
1996: 822–823). Niinpä *olla*-verbin aktiivin indikaatiivin preesensmuotojen *on*, *ovat* pi-  
täisi olla hyviä suomenkielisiä geneerisen suhteen koettimia. Koetinpöiminnan perus-  
teella ne eivät kuitenkaan ole produktiivisia, vaan tuottavat lähes pelkästään roskaa:  
tekstissä *pw02* verbimuoto *on* tai *ovat* esiintyy 68 kertaa, mutta vain kerran geneerisen  
suhteen koettimena:

- (70) Nämä toimijat muodostava käsitteen ”yhdistyneet palvelut” (Allied Services), joita ovat  
mm. etsintä-, pelastus- ja ympäristönsuojeluviranomaiset. (pw02: 8)

Esimerkissä 70 *ovat* ilmaisee geneeristä suhdetta, jossa yläkäsitteen suomenkielinen  
nimitys on *yhdistyneet palvelut* ja englanninkielinen nimitys *Allied Services*. Alakäsi-  
tteet ovat *etsintäviranomaiset*, *pelastusviranomaiset* ja *ympäristönsuojeluviranomaiset*.  
Virkkeessä on *olla*-verbin lisäksi kuitenkin viisi muuta mahdollista koetinta: substantii-  
vi *käsite*, lainausmerkit, sulkeet, lyhenne *mm.* ja elliptinen ilmaus *\*- ja*. Geneerisen suh-  
teen paikallistamiseen on siis muitakin keinoja kuin *olla*-verbi.

Koetinehdokaiden käsinpöiminnassa laajemmasta aineistosta *fi12* löytyi 34 tapaus-  
ta, joissa *olla*-verbi toimii geneerisen suhteen koettimena. Lähes kaikissa niissä joko  
ylä- tai alakäsitettä nimeävä termi on välittömästi ennen *olla*-verbin muotoa *on*, *ovat*,  
*voi olla*, *ollessa* tai *oli*:

- (71) Jäänmurto ja luotsaus ovat alusliikennepalveluun kiinteästi liittyviä palveluita. (mv7: 23)

Esimerkissä 71 alakäsitteen nimitys *luotsaus* on välittömästi ennen koetinehdokasta  
*ovat*, mutta yläkäsitteen nimitys *palvelu* on kolmen sanan päässä koetinehdokkaan jäl-  
keen. Välittömästi ennen koetinehdokasta onkin tavallisesti alakäsitettä nimeävä termi,  
vain yhdessä tapauksessa neljästä edeltävä termi on yläkäsitteen nimitys, kuten *komen-  
tosiltatehtävä* esimerkissä 72:

- (72) Keskeisin komentosiltatehtävä on aluksen navigointi. (kiv97: 2)

*Olla*-verbin ja sen jälkeen tulevan termin välissä on yleensä sanoja, koetinkorpuksen  
käsinpöimituissa tapauksissa jopa 10 sanaa. Jos koetinhaku tehdään korpusohjelman  
konkordanssitoiminnolla, jää kauas verbistä sijoittuva termi tulostusalueen ulkopuolelle:

- (73) Integroidut navigointijärjestelmät ovat tyypillisesti useista osista koostuvia, usein pitkän  
ajan kuluessa hankittuja kokonaisuuksia. (kiv97: 46)



Esimerkissä 73 geneerisen suhteen alakäsitteen nimitys *integroitu navigointijärjestelmä* on välittömästi ennen koetinehdokasta *ovat*, mutta yläkäsitteen nimitystä *kokonaisuus* saa hakea yhdeksän sanan päästä koettimen jälkeen. Konkordanssihaun tulokset osoittavat, että *olla*-verbi on kelvallinen koetin vain, jos konkordanssihaun voi tehdä yhdessä termin kanssa, niin että hakuehdossa verbimuotoa *on* tai *ovat* edeltää termi. Myös Grinstead (2000) käytti tätä koetinta yhdessä ennalta poimittujen termien kanssa, koska ilman termiä hakutulosten määrä on loputon, kuten Grinstead (2000: 39) asian ilmaisee, ja tuloksessa on paljon roskaa (ks. myös Vivaldi & Rodriguez 2001: 32). Myös Sisseck (2005), joka tutki koettimia tanskankielisellä ravitsemusalan (nutrition) aineistolla, hyväksyi geneeristä ylä- ja alakäsitesuhdetta ilmaiseviksi koetinehdokkaiksi vain ne, jotka esiintyivät käsitteen kanssa samassa kontekstissa ja korkeintaan kolmen sanan päässä käsitettä nimeävästä termistä. Näin roskaa määrää voitiin vähentää huomattavasti. Esimerkiksi 20 000 sanan ravitsemusalan korpuksessa tanskalainen geneerisen suhteen koetinehdokas *er* (on) esiintyy 447 kertaa. Näistä 65 kertaa se esiintyy käsitteen nimityksen yhteydessä, joista edelleen 45:ssä koettimen roolissa. Näin kelvollisten osumien tarkkuus saadaan nousemaan 70 prosenttiin. (Sisseck 2005: 388.) Tariq, Manumaisupat, Al-Sayed ja Ahmad (2003: 5) eivät ole yhtä vakuuttuneita koettimen englanninkielisen vastineen *is a* käyttökelpoisuudesta. Heidän tutkimassaan oikeustieteen korpuksessa kaava *X is a Y* tuotti vain 40 toivottua osumaa 400 tutkitusta virkkeestä, joissa tämä konstruktio esiintyi, joten automaattisessa tiedonhaussa tämä koetin tuottaa vain vähän tietoa käsitesuhteista mutta melkoisesti roskaa (ks. myös Grinstead 2000: 39; Kavanagh 1995). Jos Sisseckin tutkimuksessa otetaan huomioon kaikki koetinehdokkaan *er* tuottamat osumat, tarkkuus on kymmenen prosentin luokkaa eli likimain sama kuin Tariqin ym. tutkimuksessa.

**Yhteenveto.** Koettimien käsinpoiminnan ja koetinehdokkailla tehtyjen konkordanssihaakujen tulosten perusteella produktiivisin geneerisen suhteen koetin on alistuskonjunktio *kuten*. Lisäksi yläkäsitteen nimitys on aina välittömästi sitä ennen ja alakäsitteen nimitys melkein aina välittömästi sen jälkeen. Myös pronomini *muu* on käyttökelpoinen geneerisen suhteen koetin yhdessä konjunktoiden *ja* ja *myös* kanssa. Se tuottaa vähän roskaa ja viihtyy termin seurassa. Lyhenne *mm.* yhdessä *olla*-verbin kanssa muodossa *ovat mm.* on kelvallinen koetin. Se ei ole produktiivinen, mutta ei tuota ylen määrin roskaakaan. Verbit *sisältää* ja *sisältyä* ovat pikemmin koostumussuhteen kuin geneerisen suhteen koettimia, mutta verbi *jakaa* toimii kummankin käsitesuhteen koettimena. Verbi *luokitella* on geneerisen suhteen koetin, mutta se ei ole erityisen produktiivinen ja noin puolet osumista on roskaa. Kaksoispiste, sulkeet ja *olla*-verbi ovat käyttökelpoisia koettimia vain yhdessä tunnetun termin kanssa.

#### 7.4.5 Koostumussuhteen ylä- ja alakäsitteen suomalaiset koettimet

Koostumussuhteella tarkoitan käsitesuhdetta, jossa yläkäsite on kokonaisuus ja alakäsitteet ovat tämän kokonaisuuden osia. Geneerisen käsitesuhteen ohella koostumussuhde on saanut paljon huomiota terminologisissa tutkimuksissa, niin myös aiemmissa koettimia käsitelleissä tutkimuksissa. Koostumussuhdetta osoittavia koettimia on selvitetty varsinkin englanninkielisillä aineistoilla. Tässä jaksossa selostan suomenkieliselä aineistolla tekemääni koostumussuhteen koetinhakua.

Suomenkieliselä esitelmätekstillä *pw02* tehty koetinpöiminta ja koetinehdokkaiden käsinpöiminta aineistosta *fi12* osoittivat, että mahdollisia koostumussuhteen koettimia ovat substantiivit *osa* ja *osatekijä* sekä verbit *jakaa*, *jakautua*, *kuulua*, *koostua*, *käsittää*,

*muodostaa, muodostua, sisältää ja sisältyä*. Mahdollisia koostumussuhteen välimerkki-koettimia ei käsinpoiminta tuottanut. Tunnetuilla termeillä tehdyn konkordanssihaun ja käsinpoiminnan perusteella tein aineistosta *fi12* konkordanssihaut hakuehdoilla *osa\**, *koostuu*, *muodost\**, *sisält\**, *käsittä\**, *kuulu\** ja *jaka\**. Hakujen tulokset olen koonnut taulukkoon 43.

TAULUKKO 43. Koostumussuhteen terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
koostua	<i>koostuu</i>	15	11	73	vahva
muodostaa	<i>muodostaa/muodostavat</i>	35	9	26	heikko
sisältyä, sisältää	<i>sisältyy/sisältää/sisältävät</i>	45	14	31	heikko
jakaa	<i>voidaan jakaa/jaettu</i>	20	12	60	vahva
käsittää	<i>käsittä*</i>	19	11	58	vahva
kuulua	<i>kuuluu/kuuluvat/kuuluvia</i>	39	15	38	hyvä
osa, osatekijä	<i>osatekij*/osana</i>	32	10	31	heikko

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *osa*.** Käsinpoiminnan perusteella substantiivi *osa* ja lähimerkityksinen yhdyssana *osatekijä* ovat mahdollisia koostumussuhteen koettimia. Konkordanssihaku aineistosta *fi12* hakuehdolla *osa\** tuotti 372 osumaa, joten hakuehto on aivan liian laaja. Aineistosta *fi12* tehdyn manuaalisen koetinehdokkaiden poiminnan tulosten perusteella supistin hakuehdon muotoon *osatekij\*/osana*, joka tuotti 32 osumaa. Näistä noin kolmannes eli 10 on valideja koostumussuhteen osumia.

Suurimman osan osumista tuotti hakuehdon alkuosan substantiivi *osatekijä*, joka tuotti eri taivutusmuodoissaan 21 osumaa. Näistä kuudessa koetinehdokas osoittaa koostumussuhteen ylä- ja alakäsitteen välistä suhdetta:

- (74) Perämeren ja Selkämeren talviliikenteeseen on kehittynyt 1960-luvulta alkaen järjestelmä, jota voidaan kutsua suomalais-ruotsalaiseksi talviliikennejärjestelmäksi. Järjestelmässä on kolme osatekijää: jääsäännöt ja muut viranomaisten antamat säännöt, jäänmurtajien toiminta ja kauppalaivasto. (smr02: 39)

Esimerkissä 74 koetinehdokas *osatekijä* osoittaa yläkäsitteen *suomalais-ruotsalainen talviliikennejärjestelmä* ja alakäsitteiden *jääsäännöt ja muut viranomaisten antamat säännöt, jäänmurtajien toiminta ja kauppalaivasto* välistä koostumussuhdetta. Kahdeksassa tapauksessa koetinehdokas *osatekijä* ilmaisee, että koostumussuhteen yläkäsite jakautuu osiin, mutta osien nimityksiä ei ainakaan lähikontekstissa mainita:

- (75) Merenkulun turvallisuus laajasti ottaen voidaan nähdä toisiinsa kytkeytyneinä osatekijöinä, jotka vaikuttavat toisiinsa useilla tasoilla. (smr02: 11)

Kaikissa näissä osumissa yhtä lukuun ottamatta yläkäsite on *turvallisuus* tai jokin sen geneerisistä alakäsitteistä (*merenkulun turvallisuus, merikuljetusten turvallisuus*).

Hakuehdon *osatekij\*/osana* loppuosa tuotti aineistosta kaikkiaan 11 osumaa, joista neljä eli noin kolmannes on valideja:

- (76) VTMISS käsittää meritse tapahtuvat kuljetukset osana kokonaiskuljetusketjua eikä vain meriliikenteenä. (pw02: 7)

Hakuehto *osatekijä\*/osana* tuottaakin suurelta osin muuta kuin koostumussuhteen tapauksia. Kirjoittajat näkyvät käyttävän hakuehdon substantiiveja pikemmin funktiosuhteen kuin koostumussuhteen yhteydessä. Esimerkissä 75 sanotaankin suoraan, että merenkulun turvallisuuden osatekijät vaikuttavat toisiinsa eli ovat pikemmin vaikuttavia tekijöitä kuin osia. Substantiivi *osa* on muutoinkin hankala koetinehdokas, sillä hakuehtoa on vaikea tarkentaa niin, ettei hakutulokseen tulisi roskaa, sillä produktiivisimmatkin muodot *osatekijä\** ja *osana* tuottavat vain yhden kolmanneksen valideja osumia ja kaksi kolmannesta roskaa. Koetinehdokkaan hyvä puoli on, että se ei ole kirjoittajakohdainen, sillä validit osumat ovat kuudesta eri asiakirjasta.

Manuaalisen koetinpoiminnan perusteella koostumussuhteen koettimina toimivat kuitenkin pikemmin verbit kuin substantiivit. Todennäköisiä koostumussuhteen koettimia ovat verbit *jakaa*, *jakautua*, *kuulua*, *koostua*, *käsittää*, *muodostaa*, *muodostua*, *sisältää* ja *sisältyä*. Selostan seuraavaksi näiden verbien toimintaa koostumussuhteen koettimina.

**Koetinehdokas *koostua*.** Verbistä *koostua* muodostettu hakuehto *koostuu* tuotti 15 osumaa, joista 11:ssä koetinehdokas toimii koostumussuhteen ylä- ja alakäsitteiden välisen suhteen ilmaisimena. Myös neljä kelpaamattomiksi arvioitua tapausta ovat muodollisesti kelpoisia, mutta on epävarmaa, ovatko niissä mainitut käsitteet, esimerkiksi *operointiympäristö* ja *tietokannan tilasto*, merenkulun turvallisuuskäsitteitä. Kaikista osumista kahdeksan eli noin puolet on samasta asiakirjasta, ja kahdeksan aineiston kahdestatoista asiakirjasta ei tuottanut yhtään hakuehdon täyttävää osumaa, joten tämä koetinehdokas näyttäisi olevan kirjoittajakohdainen. Yhdeksässä tapauksessa eli lähes kaikissa valideissa osumissa koostumussuhteen yläkäsitteen nimitys aloittaa päälauseen ja koetinehdokas on lauseessa välittömästi yläkäsitettä nimeävän termin jälkeen:

- (77) Liikenteenjakojärjestelmä koostuu liikenteenjakyöhykkeistä, joiden avulla vastakkaisiin suuntiin kulkevat laivat ohjataan omille liikennekaistoilleen. (smr02: 70)

Esimerkissä 77 koostumussuhteen yläkäsitteen nimitys *liikenteenjakojärjestelmä* on välittömästi ennen koetinehdokasta *koostuu* ja alakäsitteen nimitys *liikenteenjakyöhyke* välittömästi koetinehdokkaan jälkeen. Kahdessa muussa tapauksessa yläkäsitteen nimitys ja koetinehdokas ovat *että*-lauseessa, mutta näissäkin tapauksissa koetinehdokas on välittömästi yläkäsitettä nimeävän termin jälkeen. Jos kokonaisuutta nimeävä termi tunnetaan, voidaan roskan määrää rajoittaa liittämällä tunnettu termi hakuehtoon ennen koetinehdokasta. Alakäsitteitä on haettava kauempaa kontekstista, sillä niitä nimeävät termit ovat välittömästi koetinehdokkaan jälkeen vain kuudessa tapauksessa.

Verbikoetinehdokas *koostuu* saattaa liittyä toiseen koostumussuhteen koetinehdokkaan. Esimerkissä 78 toinen koetinehdokas on substantiivi *osatekijä*. Kirjoittaja saattaa myös sanoa esimerkiksi, että jokin kokonaisuus koostuu seuraavista osista, kuten esimerkissä 79:

- (78) Meriturvallisuus koostuu kolmesta osatekijästä: alusturvallisuudesta, väyläturvallisuudesta (mukaanlukien luotsaus, väylien rakenteellinen taso sekä väylämerkinnän taso), ja meripelastuksesta. (hv92: 11)

- (79) Laivan rakenneturvallisuus koostuu seuraavista osista: – ehjän laivan vakavuus eri käyttötilanteissa, – rungon osastointi ja vuotavan laivan vakavuus, - - (lat97: 17)

Kummassakin esimerkissä on eksplisiittisesti ilmaistu, että jäljempänä mainitut käsitteet ovat kokonaisuuden osia eli ennen koetinehdokasta mainitun yläkäsitteen koostumussuhteisia alakäsitteitä.

**Koetinehdokkaat *muodostaa* ja *muodostua*.** Verbien *muodostaa* ja *muodostua* kelpoisuutta koostumussuhteen koettimeksi testasin aineistosta *fi12* konkordanssihaun hakuehdolla *muodost\**, jolla osumalistaan saadaan kummankin verbin taivutusmuodot. Hakuehto tuotti 86 osumaa. Käsinoiminnan perusteella aineistossa *fi12* on 11 kontekstia, joissa jompikumpi verbeistä *muodostaa* tai *muodostua* toimii koostumussuhteen koettimena, joten hakuehdon *muodost\** tuottamassa konkordanssilistassa on huomattavan paljon roskaa. Käsinoiminnan tuottamien kontekstien lähempi tarkastelu osoittaa, että niissä verbit esiintyvät vain aktiivin indikaatiivin presensin yksikön tai monikon 3. persoonan muodoissa *muodostuu*, *muodostaa*, *muodostuvat* ja *muodostavat*. Verbin *muodostua* eri taivutusmuodot toimivat koettimena vain kahdessa käsinoiminnan tuottamassa kontekstissa, joten jos hakutulokseen halutaan mahdollisimman vähän roskaa, hakuehdon voi kirjoittaa muotoon *muodostaa/muodostavat*. Tällä hakuehdolla konkordanssihaun tuotti 35 osumaa. Näistä yhdeksässä koetinehdokas toimii koostumussuhteen koettimena. Kaikissa valideissa osumissa yhtä lukuun ottamatta joko yläkäsitettä tai alakäsitettä nimeävä termi on liimautunut koetinehdokkaaseen joko välittömästi sitä ennen tai sen jälkeen:

- (80) Laivaturvallisuus muodostaa merenkulun turvallisuuden osajoukon, jossa painopiste on itse laivassa ja sen käytössä. (lat97: 14)

Seitsemässä tapauksessa alakäsitteen nimitys on ennen koetinehdokasta ja yläkäsitteen nimitys sen jälkeen, kuten esimerkissä 80 alakäsitteen nimitys *laivaturvallisuus* ja yläkäsitteen nimitys *merenkulun turvallisuus*. Näistä tapauksista kuudessa alakäsitteen nimitys on välittömästi ennen koetinehdokasta.

Koetinehdokkaalle *muodostaa/muodostavat* on tyypillistä, että sen seurassa viihtyy aina jokin toinen koetinehdokas. Kolmessa tapauksessa tämä on substantiivi *osa* ja kuudessa tapauksessa substantiivi *kokonaisuus*, *osajoukko* tai *käsite*. Verbi *muodostaa* käytetty koostumussuhteen koettimena siis hyvin samalla tavalla kuin verbi *koostua*, joka myös usein ottaa seurakseen substantiivin *osa* tai *osatekijä*. Tyypillistä on, että alakäsitteet muodostavat kokonaisuuden tai yksi alakäsite muodostaa kokonaisuuden osan, kuten esimerkissä 81, tai kokonaisuuden osajoukon, kuten esimerkissä 82:

- (81) Vaarallisten aineiden kuljetukset muodostavat oman erityisen osansa yksikkölastiliikenteestä. (smt99: 84)

- (82) Toisaalta on todettava, että laiva-alan tutkimuksen painopiste on ollut luotettavuudessa (engl. reliability), joka muodostaa laivaturvallisuuden osajoukon, mutta siitä puuttuu laivaturvallisuuteen oleellisesti kuuluvien onnettomuusskenaarioiden tarkastelu. (lat97: 2)

Käsitteiden *vaarallisten aineiden kuljetus* ja *yksikkölastiliikenne* välillä on koostumussuhde, jos *yksikkölastiliikenne* ajatellaan toiminnaksi, joka koostuu osista. Yhtä hyvin käsitteiden voidaan ajatella olevan geneerisessä suhteessa toisiinsa, jolloin toimintakäsite *vaarallisten aineiden kuljetus* on eräänlaista *yksikkölastiliikennettä* ja näin *yksikkölastiliikenteen* geneerinen alakäsite.

**Koetinehdokkaaat *sisältyä* ja *sisältää*.** Olen aiemmin jaksossa 7.4.4 geneerisen käsitesuhteen koetinehdokaiden yhteydessä selostanut konkordanssihakua aineistosta *fi12* hakuehdolla *sisält\** ja tarkennetulla hakuehdolla *sisältyy/sisältä\**. Samassa jaksossa totesin, että verbit *sisältyä* ja *sisältää* tuottivat enemmän tapauksia koostumussuhteen ylä- ja alakäsitteen välisestä suhteesta kuin geneerisestä suhteesta, vaikka englanninkielinen vastineverbi *include* on tutkimusten mukaan nimenomaan geneerisen suhteen koetin. Konkordanssihaku aineistosta *fi12* hakuehdolla *sisältyy/sisältä\** tuotti 62 osumaa, joista 17:ssä koetinehdokas toimii koostumussuhteen koettimena. Yläkäsitettä nimeävä termi on välittömästi ennen koetinehdokasta 11 tapauksessa, kuten esimerkissä 83 termi *laivan operointi*. Kolmessa tapauksessa välittömästi ennen koetinehdokasta on pronomini *johon*, kuten esimerkissä 84:

- (83) Miehistön kannalta laivan operointi sisältää siis myös merelle lähdön ja satamaan saapumisen valmistelutoimenpiteet. (lat97: 16)
- (84) Nopeuden kasvaessa yli seitsemän solmun siirrytään intergoidun navigointijärjestelmän automaattiohjaukseen, johon sisältyy useita ohjailumetodeita eli -moodeja sekä nopeuden säätöön että suunnan pitoon. (kiv97: 49)

Alakäsitteen nimitys on välittömästi koetinehdokkaan jälkeen vain kahdeksassa tapauksessa, mutta koetinehdokkaan ja termin välissä on kuitenkin vain yksi tai kaksi sanaa. Vähiten roskaa tämä koetinehdokas tuottaa hakuehdolla *sisältää/sisältävät/sisältyy*, kun haku tehdään yhdessä koostumussuhteen yläkäsitettä nimeävän termin kanssa niin, että termi on välittömästi tai korkeintaan kahden sanan päässä ennen koetinehdokasta.

Konkordanssihaku koetinehdokkailla *sisältää* tai *sisältyä* hakuehdolla *sisältyy/sisältä\** tuottaa valideja osumia neljästä asiakirjasta aineistosta *fi12*, joka koostuu 12 asiakirjasta. Koetinehdokas on kirjoittajakohmainen, sillä valideista osumista puolet on samasta asiakirjasta (kiv97), vaikka sen osuus koko aineistosta on vain neljännes, jos mittarina käytetään asiakirjojen sanalukua.

**Koetinehdokas *jakaa*.** Jaksossa 7.4.4 tarkastelin geneerisen käsitesuhteen koetinehdokkaita ja totesin, että myös verbi *jakaa* tuottaa enemmän koostumussuhteen kuin geneerisen suhteen tapauksia, mikä vastaakin arkikäsitystä tämän verbin käytöstä. *Jakaa*-verbillä koostumussuhteen osuus valideista tapauksista on kuitenkin vain kaksi kolmasosaa, kun verbeillä *sisältää* ja *sisältyä* lähes kaikissa valideissa tapauksissa on koostumussuhde. Aineistosta *fi12* tehdyn käsinpoiminnan perusteella asetin *jakaa*-verbin konkordanssihaun hakuehdoksi *voidaan jakaa/jaettu*. Tämä hakuehto tuotti 20 osumaa, joista vain kaksi oli roskaa. Valideissa osumissa oli käsitteiden välillä joko geneerinen tai koostumussuhde. Valideista kahdestatoista koostumussuhteen osumasta yhdeksässä koettimena on sanaliitto *voidaan jakaa*:

- (85) Kommentosiltaviestintä voidaan jakaa sisäiseen ja ulossuuntautuvaan viestintään. (kiv97: 66)

Modaaliapuverbi *voida* liittyy siis kolmeen neljästä koostumussuhteen tapauksesta. Verbimuoto *jaettu* on koettimena vain kolmessa validissa osumassa:

- (86) Laivamalli on jaettu tankkeihin ja osastoihin aluksen vesitiiviiden mukaan. (smt99: 109)



Esimerkissä 86 yläkäsitteen nimityksen *laivamalli* ja koetinehdokkaan *jaettu* välissä on *olla*-verbin muoto *on*, jonka voidaan tosin ajatella kuuluvan koetinehdokkaaseen. Olen jättänyt passiivirakenteeseen kuuluvan *olla*-verbin muodon *on* hakuehdosta pois, koska sen ja verbimuodon *jaettu* välissä saattaa olla sanoja. Hakuehto *voidaan jakaa/jaettu* tuottaa tasaisesti osumia aineiston *fi12* eri asiakirjoista.

**Koetinehdokas *käsittää*.** Verbien *jakaa*, *jakautua*, *kuulua*, *koostua*, *muodostaa*, *muodostua*, *sisältää* ja *sisältyä* lisäksi verbi *käsittää* osoittautui käsinpoiminnassa mahdolliseksi koostumussuhteen koettimeksi. Hakuehdolla *käsittää\** tehty konkordanssihaku aineistosta *fi12* tuotti 19 osumaa, joista 11:ssä koetinehdokas toimii koostumussuhteen koettimena. Valideista osumista yhdeksän on samasta asiakirjasta (smr02), joten tämä koetinehdokas näyttäisi olevan kirjoittajakohtainen. Verbin *käsittää* voisi kaikissa valideissa koostumussuhteen osumissa korvata jollakin verbimuodoista *kuuluu*, *sisältää* tai *koostuu* niin, ettei lauseen sisältö muuttuisi. Ilmeisesti verbiä *käsittää* koostumussuhteen ylä- ja alakäsitteiden välisen suhteen ilmaisimena käyttävä kirjoittaja on mieltynyt tähän verbiin. Hakuehdon *käsittää\** tuottamissa koostumussuhdetapauksissa ylä- ja alakäsitteitä nimeävien termien järjestys on sama kuin hakuehtojen *sisältyy/sisältää\** ja *voidaan jakaa/jaettu*. Kaikissa muissa paitsi yhdessä tapauksessa yläkäsitettä nimeävä termi on ennen koetinehdokasta ja alakäsitettä tai alakäsitteitä nimeävät termit koetinehdokkaan jälkeen. Yläkäsitettä nimeävä termi on välittömästi ennen koetinehdokasta kahdeksassa tapauksessa, kuten esimerkissä 87 termi *merenkulun toimintaympäristö*:

- (87) Merenkulun toimintaympäristö käsittää komentosillan laitteinen (kpl 4.1 ja 4.2) ja toimintatavat komentosillalla (kpl 4.3 ja 4.4) sekä aluksen ulkopuolisen liikenneympäristön (kpl 4.6). (kiv97: 41)

**Koetinehdokas *kuulua*.** Manuaalisen poiminnan tuottamista mahdollisista koostumussuhteen verbikoettimista viimeisenä tarkastelen verbiä *kuulua*. Hakuehdolla *kuulu\** tehty konkordanssihaku aineistosta *fi12* tuotti 82 osumaa. Samasta aineistosta tehdyn manuaalisen koetinhaun perusteella tämä koetinehdokas on valideissa osumissa muodossa *kuuluu*, *kuuluvat* tai *kuuluvia*. Tarkennettu konkordanssihaku hakuehdolla *kuuluu/kuuluvat/kuuluvia* tuotti 39 osumaa, joista 15 on valideja. Näistä yhdeksässä koetinehdokas on muodossa *kuuluu*, viidessä muodossa *kuuluvat* ja yhdessä muodossa *kuuluvia*. Koostumussuhteen yläkäsitteen nimitys on ennen koetinehdokasta ja alakäsitteen tai alakäsitteiden nimitykset koetinehdokkaan jälkeen 11 validissa osumassa:

- (88) Normaalityöinnassa vakiorutiineihin kuuluvat lähtövalmistelut ja reittisuunnittelu, ohjailu- ja navigointijärjestelmien käyttö sekä käytön seuranta ja viestintärutiinit. (kiv97: 61)

Esimerkissä 88 yläkäsitteen nimitys *vakiorutiinit* on välittömästi ennen koetinehdokasta ja kolmen alakäsitteen nimitykset *lähtövalmistelut ja reittisuunnittelu*, *ohjailu- ja navigointijärjestelmien käyttö* sekä *käytön seuranta ja viestintärutiinit* välittömästi sen jälkeen. Vastaavia tapauksia, joissa ainakin jommankumman käsitteen nimitys on välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen on validien osumien joukossa 11. Neljässä tapauksessa käsitteitä nimeävien termien ja koetinehdokkaan välissä on yksi sana:

- (89) Päällikön tukena on aluksen johtoryhmä, johon kuuluvat tyypillisesti päällikkö, konepäällikkö, yliperämies ja matkustaja-aluksilla intendentti. (kiv97: 90)



Esimerkissä 89 yläkäsitteen nimityksen *aluksen johtoryhmä* ja koetinehdokkaan välissä on sivulauseen aloittava pronomini *johon* ja koetinehdokkaan ja alakäsitteiden luettelon välissä sana *tyypillisesti*.

Hakuehdon *kuuluu/kuuluvat/kuuluvia* osista produktiivisin oli tämän tutkimuksen aineistossa verbimuoto *kuuluu*, joka tuotti 15 validista osumasta yli puolet eli yhdeksän osumaa. Niistä kuudessa yläkäsitettä nimeävä termi on välittömästi ennen koetinehdokasta. Jos konkordanssihaku tehdään suuresta aineistosta, voi hakuehdon tarkentaa muotoon *kuuluu*, jolloin roskan määrä vähenee noin puoleen. Aineistosta *fi12* hakuehto *kuuluu* tuotti 21 osumaa, joista siis noin puolet eli yhdeksän on valideja. Parhaaseen tulokseen päästään kuitenkin yhdistelmähaulla, jossa konkordanssihaku tehdään tunnetun termin ja koettimen yhdistelmällä. Tämä koetinehdokas ei ilmeisesti ole kirjoittajakohainen, sillä konkordanssihaku tuotti valideja osumia kuudesta asiakirjasta.

**Yhteenveto.** Käsinpoiminnan ja koetinehdokkailla tehtyjen konkordanssihakujen perusteella koostumussuhteen koettimista vahvimpia ovat verbit *jakaa* ja *käsittää*. Ne tuottavat kohtuullisen määrän osumia, joista yli puolet on valideja. Verbiin *jakaa* liittyy tämän tutkimuksen aineistossa kolmessa tapauksessa neljästä modaaliapuverbi *voida* silloin, kun verbi toimii koettimena. Verbin *koostua* tuottamista osumista noin 2/3 on valideja mutta tämä koetin on kirjoittajakohtainen samoin kuin verbi *käsittää*. Verbien *koostua*, *jakaa* ja *käsittää* hyvä puoli on, että ne liittyvät kiinteästi ainakin yhteen termiin. Tämän tutkimuksen aineistossa verbikoetin *koostua* on puhtaasti koostumussuhteen koetin, toisin kuin sen englanninkielinen vastine *consist(s) of*, joka on Pearsonin (1998: 146–147) mukaan myös geneerisen suhteen koetin. Verbit *kuulua*, *muodostaa*, *sisältyä* ja *sisältää* tuottavat runsaasti osumia mutta suuri osa niistä on kelvottomia. Verbi *muodostaa* näyttäisi tyypillisesti tarvitsevan vahvistukseksi substantiivikoettimen *osa*, *kokonaisuus*, *osajoukko* tai *käsite*. Koostumussuhteen verbikoettimet tuottavat parhaan tuloksen, kun haku tehdään yhdessä tunnetun termin kanssa. Hakua on hyvä kokeilla kaikilla verbikoettimilla, sillä osa niistä on kirjoittajakohtaisia. Kaikille koostumussuhteen verbikoettimille on yhteistä, että yläkäsitteen nimitys on ennen koetinta ja alakäsitteen nimitys tai nimitykset sen jälkeen.

Substantiivikoetin *osa* ja siitä muodostettu yhdyssana *osatekijä* tuottavat runsaasti osumia mutta myös roskaa. Niistä yleisemmän, substantiivin *osa*, hyvä puoli on, että se ei ole kirjoittajakohtainen.

#### 7.4.6 Vieruskäsitteiden suomalaiset koettimet

Vieruskäsitteet ovat hierarkkisen käsitejärjestelmän saman käsitetason käsitteitä. Ne voivat olla geneerisen, koostumussuhteisen tai geneerisiä suhteita ja koostumussuhteita sisältävän sekakoosteisen käsitejärjestelmän käsitteitä, mutta oleellista on, että ne ovat samassa käsitejärjestelmässä samalla abstraktiotasolla. Tekstissä vieruskäsitteet annetaan yleensä luettelona, joka on samalla hierarkkisen käsitejärjestelmän alakäsitteiden luettelo. Yläkäsitteen nimitys löytyy sekin samasta kontekstista. Joitakin mahdollisia vieruskäsitteiden koettimia olenkin käsitellyt geneerisen suhteen ja koostumussuhteen koettimien yhteydessä jaksoissa 7.4.4 ja 7.4.5. Tässä jaksossa tarkastelen vain niitä koetinehdokkaita, jotka toimivat vieruskäsitteiden koettimina, mutta eivät ylä- ja alakäsitteiden välisen käsitesuhteen koettimina.

Suomenkielisellä esitelmätekstillä *pw02* tehdyssä koetinpoiminnassa hyviksi vieruskäsitteen koettimiksi osoittautuivat sanat *sekä*, *kuten* ja *esimerkiksi*. Muita mahdollisia vieruskäsitteiden koettimia ovat tunnetuilla termeillä tehdyn konkordanssihaun perusteella sanaliitto ja *muut* sekä lyhenne *mm.* (ks. myös Pasanen 2005a: 295). Laajemmasta

aineistosta *fi12* tehty käsinpoiminta tuotti mahdollisten vieruskäsitteiden koettimien luetteloon sanat *lisäksi, asemasta, sijasta, kuin* ja *myös* sekä sanaliitot *niin kuin* ja *toisaalta toisaalta*. Nämä koetinehdokkaat tuottivat käsinpoiminnassa kuitenkin korkeintaan kaksi tapausta, eikä niitä sen vuoksi otettu konkordanssihaussa huomioon. Koetinehdokkaita *kuten, esimerkiksi, ja muut* ja *mm.* olen käsitellyt geneerisen käsitesuhteen koettimien yhteydessä jaksossa 7.4.4, joten tässä jaksossa rajoitan tarkastelun rinnastuskonjunktioon *sekä* ja aiempien tutkimusten perusteella mahdolliseen vieruskäsitesuhteen koetimeen, konjunktioon *ja*. Tarkastelun tulokset ovat taulukossa 44.

TAULUKKO 44. Vieruskäsitteiden terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
sekä että	<i>sekä * että</i>	32	8	25	heikko
ja	<i>*- ja/ja -*</i>	301	192	64	vahva

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *sekä*.** Aineistosta *fi12* tehty konkordanssihaku hakuehdolla *sekä* tuotti 278 osumaa, mikä on moninkertainen määrä osumia manuaalisen poiminnan tuottamiin tapauksiin verrattuna, joten todennäköisesti suurin osa osumista on roskaa. Jo liki sadan osuman manuaalinen tarkastelu on työlästä, ja seuraavaksi pyrin selvittämään, voiko kelvottomien osumien määrää vähentää hakua tarkentamalla. Kun hakuehto *sekä* korvataan hakuehdolla *sekä \* että*, osumien määrä pienenee huomattavasti. Tarkennettu konkordanssihaku tuotti 32 osumaa, joista kahdeksan on validia. Tämä hakuehto näyttää tuottavan konteksteja, joissa on toisensa pois sulkevia käsitteitä, kuten käsitteet *meriliikenne* ja *sisävesiliikenne* esimerkissä 90:

(90) Merenkululaitoksella onkin nykyään käytössään tietokanta, josta saadaan tietoa sekä meriliikenteen että sisävesiliikenteen havereista vuosittain. (smr02: 86)

Liki kaikissa valideissa tapauksissa jompikumpi koetinehdokkaan *sekä \* että* rinnastamista käsitteistä on elliptinen ilmaus. Elliptisiä ovat yhtä lukuun ottamatta tapaukset, joissa termit ovat sanaliittoja tai yhdyssanoja. Esimerkissä 90 sana *haveri* on yhteinen terminosa, joka on jätetty pois ensimmäisestä sanaliittotermistä *meriliikenteen haveri*. Esimerkissä 91 on kolme termiä, joista kahden viimeisen yhteinen terminosa *toiminta* on jätetty pois yhdyssanatermistä *luotsaustoiminta*:

(91) Näin palvelutarjontaa pystytään kehittämään vastaamaan sekä alusliikenteen että luotsaus- ja jäänmurtotoiminnan vaatimuksiin. (mv4: 25)

Hakuehtoa on vaikea tarkentaa enempää, ja roskaa on joka tapauksessa runsaasti. Koetinehdokas *sekä* näyttäisi olevan kelvollinen koetin vain, jos ainakin osa tekstin termeistä on tunnettuja ja haussa *sekä*-sanaa käytetään yhdessä termin kanssa, sillä usein termit ovat sanan *sekä* välittömässä läheisyydessä. Toisin sanoen, jos kaksi termiä on yhdistetty sanoilla *sekä* tai *sekä että*, ne ovat ilmeisesti vieruskäsitteitä. Yhteinen, usein toisesta termistä puuttuva terminosa vahvistaa käsitystä, että rinnastuskonjunktio *sekä* -

että rinnastaa vieruskäsitteitä, sillä ihannetapauksessa vieruskäsitteiden nimityksissä on yläkäsitteen nimitys yhteisenä terminosana.

**Koetinehdokas *ja*.** Aiempien tutkimusten perusteella (ks. esim. Nuopponen 1993: 110) myös *ja*-rinnastuskonjunktio voisi olla vieruskäsitesuhteen koetin. Aineistossani se osoittautui kuitenkin lähinnä roskaa tuottavaksi koetinehdokkaaksi, sillä se rinnastaa usein yleiskielen sanoja eikä erikoisalan käsitteitä nimeäviä termejä. Eräs rinnastuskonjunktio *ja* käyttöön liittyvä piirre auttaa kuitenkin rajaamaan hakuehtoa niin, että konkordanssilistaan ei tulostu ylen määrin roskaa: suuressa osassa valideista osumista koetinehdokasta edeltävä sana päättyy yhdysmerkkiin tai sitä seuraava sana alkaa yhdysmerkillä. 2 366 sanan esitelmätekstissä *pw02* sana *ja* esiintyy 121 kertaa, joista 27 kertaa se on koettimen roolissa. Näistä 13 tapauksessa yhdysmerkki on joko edeltävän sanan lopussa tai seuraavan sanan alussa, joten tässä näyttäisi olevan keino vähentää roskaa määrää. Hakuehdon tarkentaminen muotoon *\*- ja/ja -\** on tarpeen, sillä aineistosta *f112* tarkentamaton hakuehto tuottaa 2 984 osumaa. Osumien määrä pienenee huomattavasti, kun hakuehtoon lisätään edeltävään tai seuraavaan sanaan kuuluva yhdysmerkki. Hakuehto *\*- ja*, jossa edeltävä sana päättyy yhdysmerkkiin, tuottaa 254 osumaa, joista yli puolet eli 168 on valideja:

- (92) Konventionaalisella simulaattorilla tarkoitetaan tässä simulaattoria, jossa on visuaalijärjestelmä ja tavallisimmat navigointi- ja hallintalaitteet. (kiv97: 104)

Vaikka osumien määrää voidaan näin vähentää alle kymmenesosaan, ei yli 200 osuman manuaalien käsittely ole mielekäästä. Koetinehdokkaaseen *ja* pätee kuitenkin sama ilmiö kuin synonymian koetinehdokkaaseen sanaan *eli*: kaikissa valideissa tapauksissa ehdokkaan kummallakin puolella on termi tai ainakin termin osa. Jos osa tekstin termeistä on tunnettuja, saadaan koetinehdokkaan *\*- ja* sekä yhden tunnetun termin yhdistelmähaulla poimittua tekstistä vieruskäsitteiden nimityksiä. Muissa osumissa ei ehdokasta edeltävä sana eikä sitä seuraava sana ole termejä, vaan yleiskielen sanoja, kuten esimerkiksi ilmauksessa *tehtävä- ja vastuualue*, tai esimerkiksi ministeriöiden nimiä, kuten *kauppa- ja teollisuusministeriö*.

Hakuehto *ja -\**, jossa seuraava sana alkaa yhdysmerkillä, tuotti 47 osumaa eli huomattavasti vähemmän kuin hakuehto *\*- ja*. Näinhän täytyy ollakin, sillä vieruskäsitteiden nimityksissä yhteinen terminosa on usein termin loppuosa eikä termin alkuosa. Yhteinen termin alkuosa onkin paremmin merkki lähikäsitesuhteesta kuin vieruskäsitesuhteesta, kuten esimerkissä 93, jossa kahden termin määriteosassa on yhteinen osa *kaksois-*:

- (93) Kaksoispohjaiselle *ja* -runkoiselle alukselle lastivuoto tapahtuisi samalla aikavälillä neljän vuoden välein. (smr02: 23)

Noin puolet eli 24 hakuehdon *ja -\** tuottamista osumista on valideja tapauksia vieruskäsitteitä nimeävien termien välisestä rinnastuksesta. Kelpaamattomissa osumissa koetinehdokkaan molemmiin puoliin on yleiskielen sana, kuten esimerkissä 94:

- (94) Rajavartiolaitys ylläpitää toimivan ja tilanteen vaihteluihin herkästi sopeutuvan meripelastuksen johtamisjärjestelmän sekä osallistuu pääsuorittajana etenkin erityisosaamista *ja* -kalustoa edellyttäviin meripelastustehtäviin. (smr02: 31)

**Yhteenveto.** Konkordanssihakujen tulosten perusteella voidaan todeta, että jos konjunktio *ja* edeltävä sana päättyy yhdysmerkkiin tai sitä seuraava sana alkaa yhdysmerkillä

ja välittömästi ennen konjunktioita tai välittömästi sen jälkeen on termi, konjunktio toimii vieruskäsitteiden koettimena. Tällöin ainakin yhden käsitteen nimitys on elliptinen. Konjunktio *sekä* toimii koettimena, jos konkordanssihaun hakuehdon voi tarkentaa termin avulla.

#### 7.4.7 Funktiosuhteiden suomalaiset koettimet

Olen edellä selostanut, miten teksteistä voi poimia käsitteiden nimityksiä, määritelmiä ja vieruskäsitteiden nimityksiä terminologisten koettimien avulla. Tarkastelin myös niitä koettimia, jotka osoittavat käsitteiden välisiä geneerisiä suhteita tai koostumussuhteita. Käsitteiden välillä voi kuitenkin olla monia muita suhteita, joita terminologiassa on tapana nimittää funktiosuhteiksi standardin ISO 704 mukaisesti. Jaksossa 7.1.1 totesin, että tämä perinteinen käsitesuhteiden luokittelu sisältö-, koostumus- tai funktiosuhteisiin on liian karkea tutkimukseni tarpeisiin. Samassa jaksossa selostin muiden tutkijoiden käsitesuhdeluokitteluja ja totesin, että käytän tutkimuksessani ISO 704 -standardin mukaisen luokittelun ja Nuopposen (1994: 86–87, 135–139) esittämän funktiosuhteiden luokittelun välimuotoa, joka tarkkuudeltaan sopii tutkimukseeni. Nuopponen (2005: 130, 133) jakaa vaikutussuhteet (influence relations) neljään alaluokkaan: puhtaat syy-seuraussuhteet (causal relations), ”kehityssuhteet” (development relations), toimintasuhteet (functional relations) ja vuorovaikutussuhteet (interactional relations). Tässä työssä jaan vaikutussuhteet Nuopposen jaottelusta poiketan vain kolmeen ryhmään: kausaalisuhteisiin, vuorovaikutussuhteisiin ja toimintasuhteisiin. Kausaalisuhteisiin kuuluvat puhtaat syy-seuraussuhteet sekä toiminnan tarkoitusta ja tulosta ilmaisevat suhteet. Vuorovaikutussuhteessa käsitteiden välisessä suhteessa on jonkinlainen syy- tai seuraustekijä (causal element), mutta kumpikaan käsite ei varsinaisesti ole toisen käsitteen syy tai seuraus (ks. Nuopponen 2005: 133). Toimintasuhteisiin kuuluvat toiminnan ja toimijan tai toiminnan kohteen väliset suhteet sekä välinesuhde, johon olen lukenut toiminnan ja toimintaan käytetyn välineen lisäksi toiminnan ja toiminnan menetelmän tai keinon väliset suhteet. Tämän jaottelun mukaisesti aloitan funktiosuhteisiin liittyvien koettimien tarkastelun kausaalisuhteista.

##### 7.4.7.1 Kausaalisuhteiden suomalaiset koettimet

Nuopponen (1994: 97–98) sekä Madsen, Pedersen ja Thomsen (2001: 19) jakavat kausaalisuhteet kahteen alaluokkaan, jotka ovat agenttikausaalisuhde (agentkausalitet, agent-result) ja tapahtumakausaalisuhde (händelsekausalitet, activity-result). Agenttikausaalisuhteessa toiminnalla on inhimillinen toimija, tietoinen tarkoitus ja lopputulos (Pilke 2000: 125, 128). Toiminnan suorittavalla toimijalla on aina jokin tarkoitus (Pilke 2000: 117), mutta toiminnan tulos voi olla toivottu tai ei-toivottu. Esimerkiksi ympäristövahinko saattaa olla ihmisen toiminnan tulos, mutta tuskin kuitenkaan toivottu. Tällaisissa tapauksissa on luontevampaa puhua toiminnan seurauksesta kuin tuloksesta. Tapahtumakausaalisuhteessa tapahtumalla on agentin ja tietoisien tarkoituksen sijasta filosofian näkemyksen mukaan syy ja seuraus. Syyllä on voima saada aikaan muutos, ja kokemuksen mukaan syy voi olla olio, olosuhde, toiminta, tila tai toinen tapahtuma. (Pilke 2000: 116; ks. myös Nuopponen 2008: 18–19.) Seurauskin voi olla esimerkiksi toiminta tai tila.

Luokittelut ovat joka tapauksessa yleistäviä, sillä Nuopposen (2005: 134) mukaan kausaalisuhteista muodostuu monimutkaisia käsitejärjestelmiä. Syyt ja seuraukset voivat olla moninaisia, syitä voi olla useita, ne voivat vaikuttaa yhdessä tai vaihtoehtoisesti. Seurauksetkin voivat olla vaihtoehtoisia tai niitä voi olla useita. Nuopposen (1994: 192)

mukaan terminologisessa analyysissä syy- ja seurauskäsitteiksi voidaan hyväksyä hyvin laajasti erilaisia käsitteitä. Tutkimuksessani tarkastelen kieltä pintarakenteen tasolla ottamatta huomioon sen syvärakennetta. Käsittelenkin erilaisia kausaalisuhteita samassa jaksossa, koska käytännössä suhteiden erottaminen on hankalaa ilman syvällistä käsiteanalyysia.

On makuasia, käsitelläänkö tarkoitusta, tulosta, syytä ja seurausta käsitepiirteinä vai käsitesuhteina. Esimerkiksi Pilke (2000: 195) on käsitellyt tarkoitusta (avsikt) määrittävänä käsitepiirteinä (det definierande kännetecknet), joka esiintyy hänen tarkastelemissa alojen, lääketieteen, juridiikan ja tekniikan, määritelmässä. Tarkoitus onkin tärkeä käsitepiirre toimintakäsitteiden määrittelyssä, sillä noin joka kolmannessa lääketieteen, joka viidennessä tekniikan ja joka kuudennessa juridiikan toimintakäsitteen määritelmässä on Pilkkeen (2000: 195) tutkimuksen mukaan mainittu toiminnan tarkoitus. Terminologiassa toiminnan ja tarkoituksen välinen käsitesuhde on toistaiseksi saanut vähän huomiota osakseen, koska se kuuluu funktiosuhteiden lukuisaan ja sekalaiseen joukkoon. Mutta esimerkiksi Madsen ym. (2001: 7) ovat jo lisänneet tarkoituksen terminologisten käsitesuhteiden listaan. Heidän esittämässään semanttisten suhteiden hierarkiassa tarkoitus kuuluu jopa samalle tasolle sisältö- ja koostumussuhteiden kanssa (ks. Madsen ym. 2001: 17). Vastaavasti toiminnan ja sen tuloksen sekä tapahtuman ja sen syyn tai seurauksen välinen käsitesuhde voidaan lisätä käsitesuhteiden luetteloon.

Syy, seuraus, tarkoitus ja tulos liittyvät usein dynaamiseen käsitteeseen, toimintaan tai tapahtumaan. On siis luonnollista, että käsitteisiin liittyvät koettimet ovat usein dynaamisia verbejä. Aineistolla *pw02* tehdyssä kokeessa havaitsin, että syyn ja seurauksen mahdollisia suomalaisia koettimia ovat verbit *johtua* ja *aiheuttaa* (Pasanen 2005a: 295) ja tarkoituksen mahdollisia koettimia ovat verbi *pyrkii* sekä substantiivit *tarkoitus* ja *tavoite*. Käsinpoiminta laajemmasta aineistosta *fi12* tuotti lisäksi mahdolliset tapahtuman syyn ja seurauksen tai toiminnan tarkoituksen ja tuloksen koettimet verbit *aiheutua*, *seurata*, *tuottaa* ja *johtaa* sekä substantiivit *syy*, *aiheuttaja*, *tulos*, *seuraus* ja *tehtävä*. Pilottikokeen ja käsinpoiminnan tulosten vahvistamiseksi tein aineistosta *fi12* konkordanssihaut hakuehdoilla *aiheutta\**, *aiheutu\**, *johtu\*/johta\**, *pyrkii/pyrkivät/pyritään*, *syy\*/syitä*, *aiheuttaja\*/aiheuttaji\**, *\*tulos/tuloksena*, *seurauksena*, *\*tavoit\**, *tarkoitukse-na on/tarkoitus on/tarkoitettu\** ja *tehtävänä on*. Hakujen tulokset olen koonnut taulukkoon 45.

TAULUKKO 45. Kausaalisuhteen terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
aiheuttaa	<i>aiheutta*</i>	118	98	83	vahva
aiheutua	<i>aiheutu*</i>	48	33	69	vahva
johtaa	<i>johtaa</i>	32	22	69	vahva
	<i>johtava*</i>	15	4	27	heikko
johtua	<i>johtuva*</i>	10	8	80	hyvä
	<i>johtuen</i>	19	6	32	heikko
pyrkii	<i>pyrkii/pyrkivät</i>	22	15	68	vahva
syy	<i>syy/syyt/syyn/syynä/syyksi/syitä/syyhyn</i>	38	22	58	vahva
seuraus	<i>seurauksena</i>	32	17	53	vahva
aiheuttaja	<i>aiheuttaja*/aiheuttaji*</i>	5	4	80	hyvä
tarkoitus	<i>tarkoituksena on/tarkoitus on/tarkoitettu*</i>	36	20	56	vahva
tavoite	<i>*tavoitteena/*tavoite on/*tavoitteita/tavoitteina</i>	41	22	54	vahva
tulos	<i>*tulos/tuloksena</i>	21	4	19	heikko
tehtävä	<i>*tehtävänä on/tehtävä on</i>	13	13	100	vahva
johdosta	<i>johdosta</i>	9	5	56	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *aiheuttaa*.** Terminpoimintakokeen tekstillä tehdyssä koettimien poiminnassa ilmeni, että verbi *aiheuttaa* saattaa olla hyvä syyn ja seurauksen terminologinen koetin. Konkordanssihakua aineistosta *fi12* hakuehdolla *aiheutta\** tuottikin 118 osumaa, joissa kaikissa koetinehdokas ilmaisee syyn ja seurauksen välistä käsitesuhdetta. 20 osumaa on kuitenkin jätetty pois hyväksytyjen listasta, sillä niissä syy ja seuraus eivät ole merenkulun turvallisuuskäsitteitä.

Syykäsitteen nimitys on tekstissä ennen koetinehdokasta ja seurauskäsitteen nimitys koetinehdokkaan jälkeen noin 90 prosentissa osumista. Esimerkkejä päinvastaisesta järjestyksestäkin kuitenkin on. Sekä syykäsitteen että seurauskäsitteen nimitys liittyy kiinteästi koetinehdokkaaseen *aiheutta\**, sillä syykäsitteen nimitys on välittömästi ennen koetinehdokasta ja seurauskäsitteen nimitys välittömästi koetinehdokkaan jälkeen 67 prosentissa valideista osumista. Syyn ja seurauksen välillä vallitsee siis symmetria syy- ja seuraussuhdetta osoittavan koetinehdokkaan *aiheutta\** suhteen. Jos kontekstissa on syy- ja seuraussuhdetta osoittava koetinehdokas *aiheuttaa* tai *aiheuttama* eri taivutusmuodoissaan, on hyvin todennäköistä, että syy on mainittu välittömästi ennen koetinehdokasta ja seuraus välittömästi sen jälkeen, kuten esimerkissä 95:

(95) Suomen olosuhteisiin soveltuisi ns. Tuck & Taylorin menetelmä matalikon aiheuttaman imun (squat) laskemiseksi (kiv97: 82)

Esimerkissä 95 syykäsitteen nimitys *matalikko* on välittömästi ennen koetinehdokasta *aiheuttama* ja seurauskäsitteen nimitys *imu* välittömästi koetinehdokkaan jälkeen. Tässä esimerkissä yksi syy, matalikko, aiheuttaa tapahtuman imu. Mutta syitä voi olla useita, niin kuin seurauksiakin. Seuraus voi olla syyketjun aiheuttama, tai syy voi aiheuttaa



seurausten ketjun. Hakuehto *aiheutta*\* tuottaa aineistosta kahdeksan tapausta, joissa syitä on vähintään kaksi:

- (96) Jäät rikkoutuvat avoimella merialueella ja väylämerkit katoavat aaltojen ja jäiden aiheuttamiin häiriöihin. (kiv97: 83)
- (97) - - joiden tahalliset tai vakavista laiminlyönneistä johtuvat virheelliset teot tai huolimattomuus ovat todennäköisesti osaltaan olleet aiheuttamassa öljyvahinkoa. (smr02: 91)

Esimerkissä 96 tapahtuma on häiriö, jolla on kaksi syytä, aallot ja jäät. Kontekstistaan irrotettuna tämä virke on vaikeasti ymmärrettävä ja vaatii paluuta edeltävään kontekstiin. Kirjoittajan tarkoittama tapahtuma on tutkahäiriö, mutta hän käyttää siitä elliptistä ilmausta *häiriö*, sillä tekstissä on aiemmin puhuttu tutkanavigoinnista. Syyt voivat olla vaihtoehtoisia, kuten esimerkissä 97 inhimillisen tekijän vaativat toiminnat virheellinen teko ja huolimattomuus. Esimerkissä mainittujen syiden lisäksi tapahtumalla öljyvahinko voi olla muitakin syitä, kuten esimerkistä ilmenee, mutta näitä ei kirjoittaja luettele tässä kontekstissa. Vastaavasti seurauksia voi olla vähintään kaksi tai ne voivat olla vaihtoehtoisia. Esimerkissä 98 on kaksi mahdollista seurausta, tapahtumat ihmishengen menetys ja vakava ympäristövahinko, joilla on kaksi syytä, tapahtumat karilleajo ja yhteentörmäys:

- (98) Karilleajo tai yhteentörmäys voisi aiheuttaa ihmishenkien menetyksiä tai vakavia ympäristövahinkoja Suomenlahden herkälle meriluonnolle. (mv8: 9)

Esimerkissä 99 tapahtumaketju kompassitiedon häviäminen ja sitä seuraava tutkakuvan kääntyminen 'keula ylös' ovat syynä hätätilanteeseen:

- (99) Kompassitiedon häviäminen heittää tutkakuvan 'keula ylös' ja se aiheuttaa aina jonkinasteisen hätätilanteen. (kiv97: 48)
- (100) Viat ja niihin tässä luettavat inhimilliset tekijät (esim. navigointivirhe) voivat aiheuttaa laivan ohjailukyvyn menetyksen ja siitä välittömästi seuraavan karilleajon tai yhteentörmäyksen. (smr02: 43)

Esimerkissä 100 on mainittu kolme hierarkkiselta tasoltaan erilaista syytä, joista mikä tahansa voi aiheuttaa seurauksketjun. Geneerisen käsitesuhdeketjun ylimmällä hierarkiatasolla on vika, sen alapuolella inhimillinen tekijä ja alimpana inhimillisen toimijan pannota vaativa toiminta navigointivirhe, jonka mahdollisesti aiheuttama tapahtuma on laivan ohjailukyvyn menetys, jota seuraa kaksi vaihtoehtoista tapahtumaa, karilleajo tai yhteentörmäys.

- (101) Kaikki tutkan näyttölaitteiden aiheuttamat ongelmat voitaisiin ratkaista siten, että tutka näyttäisi kompassilla vakautetun todellisen liikkeen "suunta ylös" (kuva 16). (kiv97: 54)

Esimerkissä 101 olio tutkan näyttölaite on syy, joka aiheuttaa ongelman. Mahdollista on myös, että esimerkit antavat merenkulun maailmasta hiukan vääristyneen kuvan: tutkan näyttölaite ei ehkä sittenkään aiheuta ongelmaa ilman inhimillistä toimijaa, vaan kahden oliokäsitteen väliin tarvitaan jokin toiminta- tai tapahtumakäsite, jota kirjoittaja ei ole maininnut.

Kuten esimerkit osoittavat, verbikoetinehdokas *aiheuttaa* toimii koettimena erilaisen käsitteiden välisille kausaalisille käsitesuhteille. Syykäsite on tavallisimmin olio, kuten esimerkeissä 95 ja 96, toiminta, kuten esimerkeissä 97 ja 100, tai toinen tapahtu-

ma, kuten esimerkeissä 98 ja 99. Aineiston sisältöä kuvannee, että hakuehdon *aiheutta*\* tuottamissa osumissa yleisin syy on vika.

**Koetinehdokas *aiheutua*.** Verbiin *aiheuttaa* semanttisesti liittyvä verbi *aiheutua* on koettimien käsinpoiminnan perusteella mahdollinen tapahtuman syyn ja seurauksen koetin. Hakuehdolla *aiheutu*\* aineistosta *fi12* tehty konkordanssihaku tuotti 48 osumaa, joista 33 on valideja tapauksia kahden tai useamman merenkulun turvallisuuskäsitteen välisestä syy- ja seuraussuhteesta. Epäkelvoissa osumissa ei ollut syytä tai seurausta tai niissä mainitut käsitteet eivät olleet merenkulun turvallisuuskäsitteitä.

Syykäsitteen nimitys on tekstissä ennen koetinehdokasta ja seurauskäsitteen nimitys koetinehdokkaan jälkeen 28 tapauksessa eli noin 85 prosentissa valideista osumista. Näistä 22 osumassa eli 67 prosentissa kaikista valideista osumista syykäsitteen nimitys on tekstissä välittömästi ennen koetinehdokasta ja seurauskäsitteen nimitys välittömästi koetinehdokkaan jälkeen.

Aivan samoin kuin koetinehdokkaan *aiheutta*\* tuottamissa osumissa myös koetinehdokkaan *aiheutu*\* tuottamissa osumissa syitä ja seurauksia voi olla useita. Kaikkia syitä tai kaikkia seurauksia ei kuitenkaan ole välttämättä mainittu:

- (102) Suuri osa merionnettomuuksista aiheutuu laivahenkilöstön virhetoiminnasta tai laiminlyönneistä. (ato98: 1)
- (103) Maatalouden lannoitevalumista ja asutuskeskusten jätevesistä aiheutuvan rehevöitymisen lisäksi Itämeren uhkaavat öljypäästöt, - - (sms02: 36)

Esimerkissä 102 tapahtumalle merionnettomuus annetaan kaksi vaihtoehtoista syytä, toiminnot laivahenkilöstön virhetoiminta tai laivahenkilöstön laiminlyönti. Nämä syyt selittävät kuitenkin vain osan merionnettomuuksista. Esimerkissä 103 tapahtumalle rehevöityminen annetaan kaksi yhdessä vaikuttavaa syytä, oliot maatalouden lannoitevalumat ja asutuskeskusten jätevedet.

Kuten edellä annetuista esimerkeistä näkyy, koetinehdokkaan *aiheutu*\* osoittama syy- ja seuraussuhde voi vallita erilaisten käsitteiden välillä. Yhdessä kolmasosassa tapauksista syykäsite on olio ja yhtä usein se on toiminta, mutta syy voi olla myös tapahtuma tai olosuhde. Seuraus voi olla tapahtuma tai vieläkin useammin olosuhde, tila tai tilanne. Toiminta- ja tapahtumakäsitteiden välinen ero on selkeä. Syy on huomattavan usein toiminta, mutta seuraus ei koskaan.

Koetinehdokkaan *aiheutu*\* yhteydessä esiintyvä syy on seitsemässä tapauksessa jonkinlainen tekijä, joko *inhimillinen tekijä*, *ulkoinen tekijä*, *tekninen tekijä* tai *uhkatekijä*. Seuraus on yleensä tapahtumakäsite *haveri* tai *onnettomuus*.

**Koetinehdokkaat *johtua* ja *johtaa*.** Tutkimusaineiston manuaalinen tarkastelu osoitti, että verbien *aiheuttaa* ja *aiheutua* lisäksi myös verbit *johtua* ja *johtaa* ovat tapahtuman syyn ja seurauksen tai toiminnan tarkoituksen ja tuloksen mahdollisia verbikoettimia. Näillä koetinehdokkailla tein konkordanssihaun aineistosta *fi12* ensin hakuehdolla *joh-tu*\*/*johta*\*. Haku tuotti 138 osumaa. Näin monen osuman manuaalinen tarkastelu ei ole mielekäästä, joten hakuehtoa oli tarkennettava. Tarkennetun haun tein hakuehdoilla *joh-taa*, *johtuva*\*, *johtava*\* ja *johtuen*. Näistä hakuehto *johtaa* tuotti 32 osumaa, joista 22 on valideja. Tulos on varsin hyvä, sillä aineiston kokoon nähden osumia on runsaasti ja roskaa vähän. Osumalistassa on vain viisi tapausta, joissa *johtaa*-verbi ei osoita syytä ja seurausta. Lisäksi viidessä tapauksessa verbi *johtaa* osoittaa kyllä syytä ja seurausta, mutta ne eivät kuulu merenkulun alalle. *Johtaa*-verbin monimerkityksisyys kuitenkin

haittaa sen käyttöä koettimena. Esimerkissä 104 verbin *johtaa* merkitys on lähellä verbiä *johdattaa* tai *saattaa*:

- (104) HELCOMin ylimääräisessä ministerikokouksessa syksyllä 2001 ehdotettiin, että kaikki lastissa olevat yksirunkoiset säiliöalukset, joille määrättäisiin tietty rajakoko (dwt) tulisi johtaa ulkomereltä satamaan saattohinauksen avulla. (smr02: 61)

Tapauksissa, joissa koetinehdokas *johtaa* ilmaisee syyn ja seurauksen välistä suhdetta, on yleensä mainittu yksi syy ja yksi seuraus:

- (105) Talviaikaan jääpeite voi myös aiheuttaa jäävaurion laivaan. Pahimmillaan tämä voi johtaa ulkolaidan repeämiseen, mutta tämän tyypin onnettomuuden tapahtumisesta on kauan. (smr02: 41)

Esimerkissä 105 olio jäävaurio on syy, jonka seuraus voi olla tapahtuma ulkolaidan repeäminen. Syy on tässä esimerkissä ilmaistu pronomiinilla *tämä*. Koetinehdokas *johtaa* tuotti yhdeksän vastaavaa osumaa, joissa syy on koetinehdokkaan yhteydessä ilmaistu pronomiinilla *tämä*, *mikä*, *ne* tai *jotka*, koska syykäsitteen nimitys on kerrottu kontekstissa aikaisemmin. Neljässä osumassa on lauserakenne (*mikä*) *johtaa siihen, että*. Tällöin syy ja seuraus on ilmaistu monisanaisesti. Syitä voi myös olla enemmän kuin yksi:

- (106) Merenkäynnissä aluksen liikkeet ja kiihtyvyydet sekä tärinä saattavat johtaa kiinnityselimien löystymiseen ja sen jälkeen ylikuormittumiseen. (smt99: 82–83)

Esimerkissä 106 tapahtumalle kiinnityselinten löystyminen on kolme yhdessä vaikuttavaa syytä, aluksen liikkeet, aluksen kiihtyvyydet ja tärinä. Nämä syyt johtavat seurauksetjuun, jossa kiinnityselinten löystymistä seuraa toinen tapahtuma, kiinnityselinten ylikuormittuminen. Seurauksiakin voi olla useampi kuin yksi:

- (107) Toteutettu riskin arviointi paljasti myös, että sellaiset onnettomuusskenaarioihin liittyvät alkutapahtumat ja vaaratilanteet, jotka voivat johtaa alusten yhteentörmäyksiin, pohjakosketuksiin ja karilleajoihin tai alusten ajamiseen päin meressä olevia esteitä ovat varsin yleisiä tarkastelun kohteena olleessa laivaliikenteessä. (smt99: 43)

Esimerkissä 107 käsitteellä *vaaratilanne* on neljä mahdollista ja vaihtoehtoista seurausta, joita edustavat tapahtumakäsitteet *alusten yhteentörmäys*, *pohjakosketus*, *karilleajo* ja *aluksen ajaminen päin meressä olevaa estettä*.

Syy on mainittu lähes aina ennen seurausta: vain yhdessä tapauksessa seuraus on mainittu ennen syytä. Syykäsitteen nimitys on yleensä välittömästi ennen koetinehdokasta, tai sen ja koetinehdokkaan välissä on verbi *voida* tai *saattaa*. Seuraukskäsitteen nimitys on tekstissä useimmiten välittömästi koetinehdokkaan jälkeen.

Koetinehdokkaan *johtaa* tuottamissa osumissa seuraus on tapahtuma kahdessa osumassa kolmesta. Tapahtumaan johtava syy voi olla esimerkiksi olio, tapahtuma tai ominaisuus.

Hakuehdolla *johtuva\** tehty konkordanssihaku aineistosta *fi12* tuotti 10 osumaa. Kaikissa osumissa on syy- ja seuraussuhde, mutta kahdessa osumassa syy ja seuraus eivät ole merenkulun turvallisuuskäsitteitä. Tämänkin koetinehdokkaan yhteydessä syitä ja seurauksia voi olla useita. Esimerkissä 108 tapahtuman öljyvahinko vaihtoehtoisia syitä ovat tapahtumat yhteentörmäys tai karilleajo:

- (108) Yhteentörmäyksestä tai karilleajosta johtuvan öljyvahingon yhteenlaskettu riski arvioitiin 0,4 vahingoksi 1000 matkaa kohti. (smr02: 29)

Syykäsitteen nimitys on mainittu välittömästi ennen koetinehdokasta viidessä tapauksessa kahdeksasta. Seurauskäsitteen nimitys on välittömästi koettimen jälkeen kuudessa tapauksessa kahdeksasta.

Kausaalisuhte, jota koetinehdokas *johtuva\** osoittaa, voi olla erilaisten käsitteiden välillä. Sekä syy että seuraus voi olla olio, toiminta tai tapahtuma, mutta tavallisin seuraus on tapahtuma samoin kuin koetinehdokkailla *aiheutta\**, *aiheutu\** ja *johtaa*. Osumien vähäisestä määrästä huolimatta koetinehdokas *johtuva\** käy hyvin syyn ja seurauksen välisen suhteen koettimeksi, koska se tuottaa vain vähän roskaa.

Konkordanssihaku aineistosta *fi12* hakuehdolla *johtava\** tuotti 15 osumaa, joista neljä on validia:

- (109) Aluksen kaatumiseen johtava vakavuuden menetys aiheuttaa merenkulussa suurimmat tuhot, jos menetyksiä tarkastellaan kuolonuhrien lukumääränä. (smt99: 108)

Valideista osumista kolme on samasta asiakirjasta, joten koetinehdokas on kirjoittaja-kohtainen. Tämän tutkimuksen aineistosta hakuehto tuottaa runsaasti roskaa, sillä yhdeksässä osumassa 15:stä koetinehdokas *johtava\** on osa sanaliittoa *satamaan johtava väylä*.

Verbin *johtua* 2. infinitiivin muodolla *johtuen* tehty konkordanssihaku aineistosta *fi12* tuotti 19 osumaa, joista kuusi on validia. Roskan suuri määrä selittyy sillä, että 13 osumassa on sanaliitto *jostakin johtuen*. Valideissa osumissa termi on välittömästi ennen koetinta tai heti sen jälkeen:

- (110) Huolenaiheena on ollut matkustaja-alusten turvallisuuden lisäksi mm. ympäristöriskit lisääntyneestä öljyalusliikenteestä johtuen. (smr02: 8)

Esimerkissä 110 koetinehdokasta *johtuen* edeltää toimintakäsite *öljyalusliikenne*, joka on esimerkin mukaan syy käsitteelle *ympäristöriski*.

Koetinehdokkaan *johtuen* tuottamien validien osumien määrä on varsin alhainen ja kaikista osumista roskaa on lähes kolme kertaa niin paljon kuin valideja osumia. Näistä syistä koetinehdokas *johtuen* toimii koettimena vain yhdessä termin kanssa.

**Koetinehdokas *pyrkii*.** Kausaalisen käsitesuhteen verbikoetinehdokkaista *pyrkii* on mahdollinen toiminnan tarkoituksen koetin. Manuaalisen koetinehdokkaiden poiminnan perusteella tein konkordanssihaun hakuehdolla *pyrkii/pyrkivät/pyritään*. Haku tuotti 45 osumaa. Tarkastelun helpottamiseksi tein konkordanssihaun uudelleen erillisillä hakuehdoilla *pyrkii*, *pyrkivät* ja *pyritään*. Hakuehto *pyrkii* tuotti 14 ja hakuehto *pyrkivät* kahdeksan osumaa. Koetinehdokas ilmaisee niissä kuitenkin pikemmin olion kuin toiminnan tarkoitusta, vaikka tarkoituskäsite onkin kaikissa tapauksissa toimintakäsite:

- (111) Vaurionhallintajärjestelmä pyrkii automatisoimaan laajan käyttöohjeiston ja laskenta-aineiston hyödyntämisen siten, että siitä onnettomuustilanteessa on maksimiapu. (js94: 10)
- (112) Useat turvallisuusjärjestelmät pyrkivät systematisoimaan tilanteita, kasvattamaan koulutuksen määrää ja luomaan tarkistuslistoja, jotta päätöksen teossa ja toiminnoissa saavutetaan turvallisia ja johdonmukaisia käytäntöjä. (smr02: 10)
- (113) Integroitu navigointi ja Vessel Traffic Service (VTS) järjestelmä pyrkivät muuttamaan merenkulkukulttuuria tulevaisuudessa. (kl98: 68)

Esimerkissä 111 vaurionhallintajärjestelmän tarkoitus on käyttöohjeiston ja laskenta-aineiston hyödyntämisen automatisointi. Esimerkissä 112 turvallisuusjärjestelmän tarkoitus on systematisoida tilanteita, kasvattaa koulutuksen määrää ja luoda tarkistuslistoja. On epätodennäköistä, että mikään turvallisuusjärjestelmä pystyisi pelkällä olemassaolollaan täyttämään tarkoituksensa, olipa se kuinka hyvä tahansa. Siihen päästään ainoastaan järjestelmän mukaisen toiminnan kautta. Vaurionhallintajärjestelmä ja turvallisuusjärjestelmä ovatkin paremmin toiminnan välineitä kuin toimintaa, jolla on tarkoitus. Järjestelmien osina voi kuitenkin olla myös toimintakäsitteitä, vaikka olen luokitellut erilaiset järjestelmät oliokäsitteiksi. Esimerkin 113 järjestelmä *Vessel Traffic Service (VTS) järjestelmä* on tällainen käsite, sillä alusliikennettä valvova ja ohjaava alusliikennepalvelu (Vessel Traffic Service, VTS) on keskeinen osa järjestelmää. Sillä ja toiminnalla integroitu navigointi on yhteinen tarkoitus: muuttaa merenkulkukulttuuria.

*Pyrkiä*-verbin passiivimuoto *pyritään* toimii koettimena eri tavalla kuin aktiivimuodot, sillä hakuehto *pyritään* tuotti 23 osumaa, joista yhdeksässä koetinehdokas toimii toiminnan tarkoituksen koettimena. Koska *pyritään* on *pyrkiä*-verbin passiivimuoto, ei konkordanssihaun tuottamissa esimerkkilauseissa yleensä mainita eksplisiittisesti oliota tai toimintaa, joka pyrkii tai jolla pyritään tavoitteeseen:

- (114) Pyritään vaikuttamaan koulutukseen siten, että hyvän merimiestaidon ja hyvän merimiestavan taso säilytetään sellaisena kuin se kansainvälisissä sopimuksissa, koodeissa ja resoluutioissa on sovittu. (hv92: 15)

Esimerkissä 114 ei mainita toimijaa, joka vaikuttaa koulutukseen, eikä toimintaa, jolla koulutukseen vaikutetaan, vaan ainoastaan tavoite, joka koulutukseen vaikuttamalla pitäisi saavuttaa. Koska näissä tapauksissa tekstissä on ilmaistu vain toinen käsitesuhteen käsitteistä, ei tapauksia ole hyväksytty valideiksi osumiksi. Joissakin tapauksissa verbimuoto *pyritään* toimii passiivisuudestaan huolimatta toiminnan ja sen tarkoituksen koettimena. Näin on esimerkiksi 115:

- (115) Seurannan avulla ja poikkeamiin puuttumisella pyritään lisäämään merenkulun turvallisuutta ja tekemään merenkulusta helpommin ennakoitavaa. (mv5: 12–13)

Esimerkki 115 voidaan tulkita niin, että ainakin poikkeamiin puuttumisen mutta mahdollisesti myös seurannan tarkoitus on lisätä merenkulun turvallisuutta ja tehdä merenkulusta helpommin ennakoitavaa.

Koetinehdokas *pyritään* tuotti aineistosta *fi12* yhdeksän validia osumaa, mutta koska yli puolet sen tuottamista osumista ei ole kelvollisia, ei sitä voi pitää hyvänä koettimena. Lisäksi käsite, jonka tarkoitukseen koetinehdokas viittaa, on joskus haettava kaukaa tekstikontekstista.

**Koetinehdokas syy.** Tämän tutkimuksen aineistosta tehty koettimien käsinpoiminta tuotti kausaalisuhteen mahdollisten verbikoettimien lisäksi substantiivikoetinehdokkaan *syy*. Konkordanssihaku aineistosta *fi12* hakuehdolla *syy\*/syitä* tuotti 108 osumaa, joista valideja on 29. Roskan määrää lisäävät osumat, joissa on hakuehtoa vastaava substantiivi *syyskuu* eri taivutusmuodoissaan tai sanaliitot *on syytä* (tutkia, muistaa, arvioida, harkita jne.) tai *tästä syystä*. Jos nämä osumat karsitaan listasta, putoaa roskan määrä lähes puoleen, mutta jäljelle jäävien 71 osuman manuaalinen tarkastelu vie silti aikaa, joten tutkin seuraavaksi, voisiko hakuehtoa tarkentaa. Valideissa tapauksissa substantiivi *syy* on muodossa *syy*, *syyt*, *syyn*, *syynä*, *syyksi*, *syitä*, *syyhyn*, *syytä* tai *syystä*. Jos luettelosta jätetään pois muodot *syytä* ja *syystä*, jotka tarkentamattoman haun perusteella tuottavat

enimmäkseen roskaa, menetetään neljä validia osumaa. Tarkennettu haku hakuehdolla *syy/syyt/syyn/syynä/syyksi/syitä/syyhyn* tuottaa 38 osumaa, joista 22 on valideja. Hakuehto *syitä/syynä/syyksi/syyn* tuottaa 20 osumaa, joista 12 on valideja. Hakua on siis vaikea tarkentaa niin, ettei osumalistaan tulisi roskaa eikä menetettäisi valideja osumia.

Substantiivin *syy* käyttöön terminologisena *syy-* ja seuraussuhteen koettimena liittyy ongelma, jota ei muilla saman suhteen koettimilla juuri esiinny. Monesta hakuehdon *syy/syyt/syyn/syynä/syyksi/syitä/syyhyn* tuottamasta osumasta ilmenee, että seuraukselle on olemassa *syy*, mutta itse syytä ei lähikontekstissa kerrota. Esimerkissä 116 kerrotaan, että merenkulun onnettomuusriskille on enemmän kuin yksi *syy*, mutta niitä ei kerrota lähikontekstissa ennen koetinehdokasta eikä sen jälkeen:

- (116) - - päämääränä oli määritellä tekijöitä, jotka parantaisivat alusturvallisuutta rannikkovesillä analysoimalla merenkulun onnettomuusriskien taustalla olevia *syitä*. Tutkimus suoritettiin vuosina 1995–1998 kymmenen eurooppalaisen merenkulun sidosryhmiä edustavan tahon yhteistyönä (Fowler & Sörgård, 2000). (smr02: 59)

Seuraus ilman syytä on mainittu 10 tapauksessa, yhdeksässä tapauksessa on mainittu sekä *syy* että seuraus ja kolmessa tapauksessa vain *syy*. Syytä ja joskus myös seurausta on etsittävä kaukaa kontekstista, jolloin konkordanssiohjelman osumalistaan tulostunut konteksti ei riitä, vaan on tutkittava alkuperäistä lähdetekstiä. Tämä on kuitenkin niin hankalaa ja hidasta, että substantiivin *syy* käyttö terminologisena koettimena ei ehkä ole mielekästä. Merkillepantavaa on, että seuraus on yhdeksässä tapauksessa onnettomuus, merionnettomuus, onnettomuusvaara tai onnettomuusriski.

**Koetinehdokas *aiheuttaja*.** Toisen mahdollisen *syy-* ja seuraussuhteen substantiivikoettimen *aiheuttaja* toimintaa testasin tekemällä konkordanssihaun aineistosta *fi12* hakuehdolla *aiheuttaja\*/aiheuttaji\**. Haku tuotti viisi osumaa, joista neljä on validia tapausta:

- (117) Toisaalta pohjakosketukset ja karilleajot, törmäykset esteisiin ja vaikutusalueeltaan rajoitetut tulipalot arvioitiin todennäköisimmäksi aluksiin kohdistuvien vahinkojen aiheuttajaksi. (smt99: 43)

Seuraus on joko vahinko, riski, öljypäästö tai häiriö, ja jokin seurauskäsitteitä nimeävistä termeistä on kaikissa esimerkeissä välittömästi ennen koetinehdokasta. Substantiivi *aiheuttaja* on toisaalta hyvä terminologinen koetin, koska tarkalla hakuehdolla se tuottaa todennäköisesti vähän roskaa, mutta toisaalta huono koetin, koska se ei ollut kovin produktiivinen ainakaan tämän tutkimuksen aineistossa.

**Koetinehdokas *tulos*.** Substantiivien *syy* ja *aiheuttaja* lisäksi koettimien käsinpaiminta tuotti kolmannenkin mahdollisen *syy-* ja seuraussuhteen tai tarkoitus- ja tulossuhteen substantiivikoettimen, substantiivin *tulos*. Konkordanssihaku aineistosta *fi12* hakuehdolla *\*tulos/tuloksena* tuotti 21 osumaa, joista neljässä on ilmaistu toiminnan tulos tai tapahtuman seuraus:

- (118) Ilmeisimmän syyn laivan laidoituksen puhkeamiseen muodostavat ulkoisen esineen tunkema oli pa se sitten toisen laivan keula tahi kari. Tuloksena on veden tunkeutuminen alukseen siinä määrin kuin sen rakenteellinen jaottelu sallii. (js94: 8)

Esimerkissä 118 seurauskäsitteen nimitys *veden tunkeutuminen alukseen* on koetinehdokkaan jälkeen ja välissä on vain yksi sana, *olla*-verbin muoto *on*, mutta sen tapahtu-



man nimitys, jonka seurauksesta kontekstissa puhutaan, on kaukana edellisessä virkkeessä. Kolmessa validissa osumassa on tapahtumaketju, joissa jälkimmäinen tapahtuma on aiemman tapahtuman seuraus. Esimerkissä 118 tapahtumaketju alkaa ulkoisen esineen tunkemasta, jonka seuraus on laivan laidoituksen puhkeaminen ja ketjun päättää veden tunkeutuminen alukseen.

Hakuehdon tuottamille roskaosumille on yhteistä, että sanoja *tulos* tai *tuloksena* edeltää usein sana *analyysi*, *prosessi* tai *tutkimus*, kuten esimerkissä 119:

(119) Kuva 32 esittää tutkimuksen tuloksena esitettyä nopeussuosituksia. (smr02: 63)

Koetinehdokas *\*tulos/tuloksena* ei ole hyvä koetin, koska se tuottaa runsaasti roskaa eikä tarkkuutta ole mahdollista parantaa edes yhdistämällä hakuun tunnettu termi.

**Koetinehdokas *seuraus*.** Konkordanssihaku aineistosta *fil2* hakuehdolla *seurauksena* tuotti 32 osumaa, joista noin puolet eli 17 on validia. Koetinehdokkaalle on tyypillistä, että joko syykäsitteen nimitys on välittömästi ennen koetinehdokasta tai seurauskäsitteen nimitys välittömästi koetinehdokkaan jälkeen. Näin on kaikissa valideissa osumissa kahta lukuun ottamatta.

Lähes kaikki kausaalisuhteen koetinehdokkaat tuottavat osumia, joissa syitä, seurauksia, tarkoituksia tai tuloksia on enemmän kuin yksi. Koetinehdokas *seurauksena* ei ole poikkeus:

(120) Ohjelmiston vauriolaskentamalleilla (Damage analysis module) voidaan edelleen arvioida karilleajon tai yhteentörmäyksen seurauksena alukselle syntyvien vaurioiden suuruutta ja sijaintia. (smr02: 54)

Esimerkissä 120 on kaksi vaihtoehtoista syytä, joiden nimitykset *karilleajo* tai *yhteentörmäys* ovat välittömästi ennen koetinta. Kymmenessä validissa osumassa sekä syy että seuraus ovat tapahtumia, ja kaikissa valideissa osumissa ainakin jompikumpi, joko syy tai seuraus, on tapahtuma:

(121) Näihin [tekniisiin riskitekijöihin] liittyy lastin siirtyminen, jään kerrostuminen talvella kansirakennelmiin, vaikeat sääolosuhteet yleensäkin ja vuoto joko huonon kunnon tai karilleajon/yhteentörmäyksen seurauksena. (smr02: 35)

Esimerkissä 121 kolmesta vaihtoehtoisesta syystä kaksi, karilleajo ja yhteentörmäys, on tapahtumia, joiden seuraus on tapahtuma vuoto. Seuraus tai syy voi olla myös olosuhde, tila tai ominaisuus. Oliokäsitteiden puuttuminen on tyypillistä koetinehdokkaan *seurauksena* tuottamille valideille osumille, sillä olio on seurauksena vain yhdessä validissa osumassa (ks. esimerkki 120). Koetinehdokkaan heikkous on, että se saattaa olla kirjoitajakohtainen, sillä kaikki validit osumat yhtä lukuun ottamatta ovat kahdesta asiakirjasta.

**Koetinehdokas *tavoite*.** Tarkoituspääte mahdollinen substantiivikoetin *tavoite* tuotti hakuehdolla *\*tavoit\** 72 osumaa, joista 22 on validia. 13 validissa osumassa koetinehdokas ilmaisee toiminnan tarkoitusta:

(122) FSA:n viimeisen vaiheen tavoitteena on luoda lisää ymmärtämystä ja esittää perusteet rationaaliselle päätöksenteolle merenkulun riskien hallinnassa. (smr02: 53)

Yhdeksässä osumassa koetinehdokas ilmaisee olion tarkoitusta, mutta näissäkin tapauksissa kyse on paremmin olion toiminnan tarkoituksesta kuin olion tarkoituksesta:

- (123) GOFREPin ensisijaisena tavoitteena onkin pienentää alusten yhteentörmäysvaaraa. (mv8: 9)
- (124) Varustamon tavoitteena on saada aikaan paras mahdollinen talouden, tekniikan ja turvallisuuden yhdistelmä. (smr02: 93)

Esimerkissä 122 puhutaan toiminnan FSA (järjestelmällinen turvallisuusanalyysi) tarkoituksesta ja esimerkeissä 123 ja 124 olioiden GOFREP (Suomenlahden kansainvälisen merialueen kattava alusten pakollinen ilmoittautumisjärjestelmä) ja varustamo tarkoituksesta.

Hakuehdon *\*tavoit\** tuottamien osumien joukossa on kymmenen osumaa, joissa kerrotaan ensin tietty tavoite ja sen jälkeen kuvataan toiminta, joka on tarpeen tähän tavoitteeseen pääsemiseksi, kuten esimerkissä 125:

- (125) Kansainvälisen merenkulkujärjestön IMO:n (International Maritime Organization) työkentelyn päätavoitteita ovat sekä merenkulun turvallisuuden kehittäminen että merellisen ympäristön suojelu. Näiden tavoitteiden saavuttamiseksi IMO luo ja pyrkii saattamaan voimaan kansainvälisesti hyväksytyjä suosituksia ja sääntöjä. (smr02: 88)

Esimerkissä 125 kerrotaan ensin, että Kansainvälisen merenkulkujärjestön toiminnan tavoite on merenkulun turvallisuuden kehittäminen ja merellisen ympäristön suojelu. Seuraavassa virkkeessä kerrotaan, että näihin tavoitteisiin IMO pyrkii luomalla ja saattamalla voimaan suosituksia ja sääntöjä. Esimerkissä 125 substantiivin *tavoite* yhteydessä käytetään verbiä *saavuttaa*. Samaa verbiä käytetään viidessä muussa hakuehdon *\*tavoit\** tuottamassa osumassa.

Koska hakuehto *\*tavoit\** tuottaa huomattavan paljon roskaa, hakua olisi hyvä saada tarkennettua. Valideissa osumissa koetinehdokas *tavoite* on muodoissa *tavoitteena*, *perustavoitteena*, *tavoite on*, *päätavoite on*, *tavoitteita*, *päätavoitteita* tai *tavoitteina on*. Kaikki validit osumat saadaan hakuehdolla *\*tavoitteena/\*tavoite on/\*tavoitteita/tavoitteina*. Hakuehdon tarkentaminen tähän muotoon vähentää roskaa huomattavasti, sillä tarkennetulla hakuehdolla osumia saadaan 41, kun tarkentamattomalla hakuehdolla osumien määrä on 72. Eniten eli 33 osumaa tuotti muoto *\*tavoitteena*. Roskan osuus on kuitenkin noin puolet tarkennetussakin haussa. Tunnetun termin lisääminen hakuehtoon vähentää kelvottomia osumia, mutta ei paranna tulosta, sillä termi liittyy välittömästi koetinehdokkaaseen vain noin joka toisessa validissa osumassa. Koetinehdokas *tavoite* on kuitenkin hyvä koetin, koska se tuottaa runsaasti tietoa toimintakäsitteiden tarkoituksesta.

**Koetinehdokas *tarkoitus*.** Konkordanssihaku hakuehdolla *tarkoituksena on/tarkoitus on/tarkoitettu\** tuotti 36 osumaa, joista yhdeksässä on kerrottu merenkulkuun kuuluvan toiminnan tarkoitus. Näistä kahdeksassa toimintakäsite on välittömästi ennen koetinehdokasta:

- (126) Alusliikennepalvelun tarkoituksena on varmistaa liikenteen turvallisuus ja ehkäistä ympäristölle aiheutuvia haittoja antamalla aluksille niiden navigointia tukevia tietoja sekä ohjaamalla liikennettä niin, että vaaratilanteita ei pääse syntymään. (mv7: 22)

Koetinehdokas tuotti yllättäen enemmän osumia, joissa on olion tarkoitus, kuin osumia, joissa on toiminnan tarkoitus. Olion tarkoitus on kuvattu 11 osumassa:

- (127) Kansainvälisen merenkulkujärjestön IMO:n mukaan liikenteenjakojärjestelmien tarkoitus on parantaa meriturvallisuutta alueilla, missä laivaliikenne on vilkasta. (smr02: 70)

Koetinehdokas *tarkoitus* tuottaa runsaasti kelvottomia osumia, mutta sen käyttökelpoisuus paranee, jos hakuehtoon ennen koetinehdokasta liitetään toimintakäsitettä nimeävä termi, jolloin lähes kaikki osumat ovat valideja.

**Koetinehdokas *tehtävä*.** Tarkoituksen mahdollinen substantiivikoetin *tehtävä* on manuaalisen poiminnan perusteella koettimena toimiessaan yleensä muodossa *tehtävänä on*, mutta se voi olla myös muodossa *päätehtävänä on* tai *tehtävä on*. Hakuehto *\*tehtävänä on/tehtävä on* tuotti 13 osumaa. Näissä kaikissa yhtä lukuun ottamatta koetinehdokas ilmaisee olion tarkoitusta, joka ilmaistaan toimintana:

- (128) Saattohinaajan tehtävänä on varmistaa laivan turvallisuus väylällä vaikeidenkin olosuhteiden tai käyttöhäiriön aikana. (smr02: 78)

Esimerkissä 128 koetinehdokas *tehtävänä on* ilmaisee saattohinaajan tarkoitusta, joka on varmistaa laivan turvallisuus väylällä. Kahta validia osumaa lukuun ottamatta oliokäsitettä nimeävä termi tai muu nimitys on välittömästi ennen koetinehdokasta, kuten esimerkissä 128. Joskus koetinehdokkaan kontekstista käy ilmi, että oliokäsitteellä voi olla muitakin tarkoituksipiirteitä kuin kontekstissa mainittu. Esimerkissä 129 tämä on osoitettu sanan *tehtävä* määritteellä, jona esimerkissä on käytetty adjektiivin *tärkeä* superlatiivia *tärkein*:

- (129) Päällikön tärkein tehtävä on huolehtia alus turvallisesti määräpäähänsä. (kiv97: 89)

Toisessa vastaavassa esimerkissä samassa tehtävässä on käytetty substantiivimääritettä *pää*, joka muodostaa pääsanan kanssa yhdyssanan *päätehtävä*:

- (130) IBNet-järjestelmän päätehtävänä on toimia päätöksenteon tukena jäänmurtotoiminnan suunnittelussa ja seurannassa. (smr02: 80)

Tämän tutkimuksen aineistosta tulostetun konkordanssihaun osumalistan perusteella ainakin merenkulun turvallisuutta käsittelevien tekstien kirjoittajat käyttävät tarkoituksen koetinta *\*tehtävänä on/tehtävä on* varsin harvoin, mutta silloin, kun sitä käytetään, se todella osoittaa tarkoitusta.

**Yhteenveto.** Syyn ja seurauksen sekä tarkoituksen ja tuloksen suomalaisista koettimista parhaita ovat verbikoettimet, sillä ne ovat produktiivisia, tuottavat tarkennetuilla hauilla vähän roskaa ja niiden valikoima on laaja. Parhaaksi suomalaiseksi syyn ja seurauksen koettimeksi osoittautui tämän tutkimuksen aineistolla verbi *aiheuttaa*, joka oli produktiivinen ja tuotti vain vähän roskaa. Toinen vahva syyn ja seurauksen verbikoetin on verbi *aiheutua*, joka ei ollut yhtä produktiivinen, mutta 69 prosenttia osumista on valideja. Verbit *aiheuttaa* ja *aiheutua* käyttäytyvät koettimena hyvin samalla tavalla. Niiden hyvä puoli on, että 67 prosentissa valideista osumista syykäsitteen nimitys on mainittu välittömästi ennen koetinehdokasta ja seurauskäsitteen nimitys välittömästi koetinehdokkaan jälkeen. Myös verbi *johtaa* osoittautui hyväksi syyn ja seurauksen koettimeksi. Verbin monimerkityksisyys on kuitenkin ongelma, joka haittaa sen käyttöä koettimena.

Koetin *johtaa* liittyy seurauskäsitteeseen yhtä kiinteästi kuin verbit *aiheuttaa* ja *aiheutua*, sillä seurauskäsitteen nimitys on välittömästi koettimen jälkeen 68 prosentissa tapauksista. Verbin *johtaa* toiminta koettimena eroaa kuitenkin verbeistä *aiheutua* ja varsinkin verbistä *aiheuttaa* siinä, että sen ja sitä edeltävän syykäsitteen nimityksen välissä on lähes joka toisessa tapauksessa modaaliapuverbi *voida* tai *saattaa*. Verbin *johtua* partisiippiimuoto *johtuva* osoittautui vahvaksi koettimeksi, sillä 80 prosenttia hakuehdon *johtuva*\* tuottamista osumista on valideja. Hakuehdon *johtuva*\* tuottamisissa valideissa osumissa seuraus ei ole enää pelkästään mahdollinen vaan väistämätön. Tarkoituksen suomalaisista koettimista parhaaksi osoittautui tutkimukseni aineistolla verbi *pyrkii*.

Substantiivikoettimet eivät ole likikään verbien veroisia syyn ja seurauksen tai tarkoituksen ja tuloksen koettimina tutkimukseni suomalaisessa aineistossa. Syyn ja seurauksen mahdollisista suomalaisista substantiivikoettimista produktiivisimmaksi osoittautui substantiivi *seuraus* hakuehdolla *seurauksena*. Osumista noin puolet on valideja. Kaikissa niissä on termi joko välittömästi ennen koetinta, välittömästi sen jälkeen tai korkeintaan yhden sanan päässä. Substantiivin *aiheuttaja* käyttöä syyn ja seurauksen koettimena puolustaa se, että lähes kaikki osumat ovat valideja, vaikka koetin ei olekaan produktiivinen. Substantiivit *syy* ja *tulos* osoittautuivat ongelmallisiksi koetinehdokkaiksi, sillä hakuehtoa on vaikea tarkentaa niin, ettei hakutulokseen tulisi kovin paljon roskaa. Toiminnan tarkoituksen mahdollisista substantiivikoettimista produktiivisimpia olivat substantiivit *tavoite* ja *tarkoitus*. Tarkennettujen hakuehtojen tuottamista osumista noin puolet on valideja. Substantiivi *tehtävä* ei ollut yhtä produktiivinen, mutta se on vahva tarkoituksen koetin, sillä kaikki hakuehdon *\*tehtävänä on/tehtävä on* tuottamat osumat ovat valideja. Myös ilmaus *johdosta* voidaan vähän roskaa tuottavana hyväksyä suomalaisten terminologisten koettimien luetteloon, vaikka se ei ollut kovin produktiivinen.

#### 7.4.7.2 Vaikuttavan tekijän suomalaiset koettimet

Tässä jaksossa tarkastelen vuorovaikutussuhdetta eli suhdetta, jossa käsite vaikuttaa toiseen käsitteeseen. Vaikutussuhteen suunta on tunnettu kontekstin perusteella. Olen erottanut vuorovaikutussuhteet omaksi ryhmäkseen. Ne eroavat nimittäin kausaalisuhteista siten, että kumpikaan käsite ei varsinaisesti edusta toisen käsitteen syytä tai seurausta, vaikka suhteessa onkin jonkinlainen kausaalinen tekijä (ks. jakso 7.4.7).

Pilke (2000) on käsitellyt vaikuttavaa tekijää dynaamisen käsitteen käsitepiirteenä. Vaikuttava tekijä on Pilkkeen tutkimuksen tulosten perusteella tapahtumakäsitteen olennainen käsitepiirre ja tapahtuman käyntiin laittava voima (ks. esim. Pilke 2000: 155, 157). Pilkkeen luokittelussa syy on vaikuttavan tekijän alaluokka eli eräänlainen vaikuttava tekijä (ks. Pilke 2000: 235). Tässä tutkimuksessa olen kuitenkin käsitellyt syykäsitteitä erillään vaikuttavasta tekijästä. Pilkkeen aineistossa vaikuttava tekijä voi tapahtuman käsitepiirteenä esiintyessään olla toiminta, tapahtuma tai tila (Pilke 2000: 158, 235). Tämän tutkimuksen aineistossa vaikuttava tekijä voi olla funktiosuhteessa muihinkin käsitteisiin kuin tapahtumiin, vaikka suhteen toinen osapuoli on usein juuri tapahtuma. Huomionarvoista on, että Pilkkeen (2000: 158) aineistossa vaikutus tapahtumaan on aina jollain lailla negatiivinen. Oman tutkimukseni aineistossa vaikutus voi kuitenkin olla myös positiivinen.

Aineistolla *pw02* tehty poiminta ei tuottanut yhtään sellaista koetinehdokasta, joka osoittaa vaikuttavaa tekijää, mutta laajemmalla aineistolla *fi12* tehdyssä koettimien käsinpoiminnassa mahdollisiksi vaikuttavan tekijän koettimiksi osoittautuivat verbit *vaikuttaa* (*vaikuttavat*, *vaikutetaan*, *vaikuttava*), *riippua*, *parantaa* ja *vähentää* sekä substantiivit (*osa*)*tekijä*, *vaikutus*, *keskeinen osa* ja *tärkeä osa*. Mahdollisia vaikuttavan

tekijän välimerkkikoettimia ei käsinpoiminta tuottanut. Koetinehdokkaiden testaamiseksi tein aineistosta *fi12* konkordanssihaut hakuehdoilla *\*vaikutta\**, *riippuu*, *parantaa*, *vähentää*, *\*tekijä\*/\*tekijö\**, *vaikutus/vaikutuksia*, *osuus/osa\**. Seuraavaksi selostan näiden konkordanssihakujen tuloksia, jotka esitän taulukossa 46.

TAULUKKO 46. Vaikuttavan tekijän terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
vaikuttaa	<i>vaikuttaa/vaikuttavat</i> <i>*vaikuttavi*/vaikutetaan/vaikuttava</i>	60 25	40 13	67 52	vahva vahva
parantaa	<i>parantaa</i>	29	16	55	vahva
vähentää	<i>vähentää</i>	52	27	52	vahva
tekijä	<i>*av* tekij*/*äv* tekij*</i> <i>tärke* tekij*/tärke* osatekij*/</i> <i>keskei* tekij*/olennai* tekij*/</i> <i>olennai* osatekij*</i>	30 9	15 8	50 89	hyvä hyvä
vaikutus	<i>vaikutus/vaikutuksia</i>	44	18	41	hyvä
osa	<i>osaltaan</i>	15	9	60	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *vaikuttaa*.** Aineistosta *fi12* tehdyn koetinehdokkaiden käsinpoiminnan perusteella verbi *vaikuttaa* on todennäköinen vaikuttavan tekijän koetin. Käsinpoiminnan tuottamissa tapauksissa verbi on useimmiten indikatiivin preesensin yksikön 3. persoonan muodossa *vaikuttaa* tai monikon 3. persoonan muodossa *vaikuttavat*, mutta se voi olla myös passiivimuodossa *vaikutetaan* tai 1. partisiipin eri taivutusmuodoissa.

Konkordanssihaku hakuehdolla *\*vaikutta\** aineistosta *fi12* tuotti 99 osumaa. Hakuehdossa on katkaisumerkki myös hakuehdon alussa, sillä joissakin konteksteissa kirjoittajat käyttävät *vaikuttaa* verbin sijasta verbiä *myötävaikuttaa*. Konkordanssihaun osumalistan tarkastelun perusteella 99 osumasta 64 tapauksessa koetinehdokas toimii vaikuttavan tekijän koettimena merenkulun turvallisuuden kontekstissa. 39 tapauksessa eli suurimmassa osassa valideista tapauksista koetinehdokas on muodossa *vaikuttaa* tai *vaikuttavat*. Muissa valideissa osumissa koetinehdokas on 1. partisiipin *vaikuttava* eri taivutusmuodoissa, useimmiten monikkomuodoissa, kuten esimerkiksi *vaikuttavia* tai *vaikuttavista*.

Verbin passiivimuoto *vaikutetaan* ei sovi hakuehtoon *\*vaikutta\**, joten jos passiivimuodon sisältävät kontekstit halutaan hakutulokseen, on hakuehtoa laajennettava tällä muodolla. Hakuehto *vaikutetaan* tuottaa vain kolme osumaa, mutta ne kaikki ovat valideja. Suurin osa valideista tapauksista saadaan konkordanssilistaan tarkennetulla hakuehdolla *vaikuttaa/vaikuttavat/\*vaikuttavi\**, jolla alkuperäisen hakuehdon *\*vaikutta\** tuottamaa roskaa määrää saadaan pienennettyä, mutta samalla menetetään joitakin valideja osumia. Jos hakuehtoa laajennetaan passiivimuodolla *vaikutetaan* ja partisiippiimuodolla *vaikuttava*, tulokseen saadaan seitsemän validia osumaa lisää ilman roskaa.

Valideissa osumissa joko vaikuttava tekijä tai sen vaikutuksen kohde on mainittu koettimen lähikontekstissa, mutta vain harvoin ne ovat välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen. Vaikuttavia tekijöitä voi olla useita, kuten esimerkissä

131, jossa väylän leveyteen vaikuttaa neljä tekijää: aluksen mitat, paikanmäärittäminen, aluksen ohjailuominaisuudet ja tuulen aiheuttama sorto:

- (131) Väylän leveyteen vaikuttavat aluksen mitat, paikanmäärittäminen tarkkuus, aluksen ohjailuominaisuudet ja tuulen aiheuttama sorto. (kiv97: 83)

Vaikutussuhde voi olla erilaisten käsitteiden välillä. Usein vaikuttava tekijä on toiminta:

- (132) Kokonaisturvallisuuden käsitteeseen vaikuttaa toki muitakin tekijöitä, kuten esimerkiksi SAR (Search and Rescue) -toimenpiteet ja satamatoiminnot (lähinnä purku ja lastaus). (smr02: 11)

Esimerkissä 132 pelastustoiminta on yksi kokonaisturvallisuuteen vaikuttava tekijä, mutta vaikutuksen laatu ei käy esimerkkikontekstista ilmi. Muissakaan valideissa osumissa ei vaikutuksen suuntaa tai laatua ole yleensä ilmaistu. Halutessaan korostaa suuntaa tai laatua kirjoittajat käyttävät esimerkiksi verbejä *parantaa* tai *heikentää* verbin *vaikuttaa* sijasta.

Konkordanssihaun tulosten perusteella koetinehdokas *vaikuttaa* tuottaa tutkimukseni aineistosta eniten valideja osumia ja vähiten roskaa hakuehdolla *vaikuttaa/vaikuttavat/\*vaikuttavi\*/vaikutetaan/vaikuttava*. *Vaikuttaa* on vahva koetin, sillä se on produktiivinen ja suurin osa osumista on valideja. Toistensa kanssa vaikutussuhteessa olevien käsitteiden nimitykset saattavat kuitenkin olla tekstissä kaukana koettimesta.

**Koetinehdokas *riippua*.** Käsinpoiminnan perusteella verbi *riippua* on toinen mahdollinen vaikuttavan tekijän verbikoetin. Käsinpoiminnan tulosten mukaan verbi *riippua* on koettimena toimiessaan aktiivin indikatiivin preesensin yksikön 3. persoonan muodossa *riippuu*. Konkordanssihaku aineistosta *fi12* hakuehdolla *riippuu* tuotti kuitenkin vain neljä osumaa. Näistä kahdessa verbimuoto toimii terminologisena koettimena:

- (133) Jäättä esiintyy kaikkina talvina, ja jääpeitteen laajuus riippuu talven ankaruudesta. (smr02: 10)

Esimerkissä 133 ominaisuus talven ankaruus on vaikuttava tekijä, joka vaikuttaa ominaisuuteen jääpeitteen laajuus. Verbi *riippua* ei siis ole kovin tuottelias koetin, mutta koska roskan määrä on ainakin tässä aineistossa vähäinen, sitä voi käyttää vaikutussuhteen koettimena.

**Koetinehdokkaat *parantaa* ja *vähentää*.** Verbit *vähentää* ja *parantaa* osoittautuivat manuaalisessa koetinhaussa hyviksi vaikuttavan tekijän koettimiksi. Konkordanssihaku hakuehdolla *parantaa* tuottikin 29 osumaa, joista 16 on valideja:

- (134) Valmis verkko parantaa oleellisesti merenkulun turvallisuutta välittämällä monenlaista tärkeää tietoa niin aluksista kuin aluksillekin. (mv10: 11)

Esimerkissä 134 vaikuttava tekijä on olio verkko, joka vaikuttaa positiivisesti merenkulun turvallisuuteen. Konkordanssihaku hakuehdolla *vähentää* tuotti sekin runsaasti osumia. 24 osumasta 20 on valideja:

- (135) Työ on tärkeää, koska riskianalyyysien mukaan ilmoittautumisjärjestelmän käyttöönotto vähentää yhteentörmäyksen vaaraa jopa 80 prosentilla. (mv9: 10)



Esimerkissä 135 vaikuttava tekijä on toiminta ilmoittautumisjärjestelmän käyttöönotto, jonka vaikutus on positiivinen, sillä se pienentää yhteentörmäyksen vaaraa. Ilmeisesti kirjoittaja on tarkoittanut, että vaikuttava tekijä on olio ilmoittautumisjärjestelmä, jonka vaikutus jatkuu toivottavasti koko sen olemassaolon ajan, eikä toiminta ilmoittautumisjärjestelmän käyttöönotto, joka on olemassa vain tietyn lyhyen ajanjakson eli järjestelmän käyttöönottovaiheen ajan.

**Koetinehdokas *tekijä*.** Käsinoiminnassa parhaaksi vaikuttavan tekijän substantiivi-koettimeksi osoittautui substantiivi *tekijä*. Kordanssihaku aineistosta *fi2* hakuehdolla *\*tekijä\*/\*tekijö\** tuotti 176 osumaa. Tarkennetulla hakuehdolla *tekijä\*/\*tekijö\*/osatekijä\** osumien määrä jäi 100 osumaan, mutta tässäkin konkordanssilistassa on todennäköisesti vielä monta roskaosumaa, sillä käsinoiminta tuotti vain 23 sellaista tapausta, jossa koetinehdokas toimii vaikuttavan tekijän koettimena. Konkordanssihaun osumalistan tarkastelu osoitti, että valideissa osumissa on substantiivien *tekijä* tai *osatekijä* yhteydessä lähes aina jokin määrite. Määritteenä voivat toimia partisiippimuodot *vähentävä*, *parantava*, *nostava*, *lisäävä*, *myötävaikuttava* tai *vaarantava*:

- (136) Jäätäminen on pohjoisten alueiden erityinen laivan painopistettä nostava ja siten vakavuuksi vaarantava tekijä. (smt99: 17)

Esimerkissä 136 koetinehdokkaan *tekijä* määritteenä on *vaarantaa*-verbin partisiippimuoto *vaarantava*. Vaikuttava tekijä on tapahtuma jäätäminen, joka vaikuttaa ominaisuuteen vakavuus. Kaikki määriteosat ovat partisiippimuotoja, joissa on partisiipin tunnus *-ava* tai *-ävä*. Näin hakuehto voidaan muokata muotoon *\*av\* tekijä\*/\*äv\* tekijä\**. Tällä hakuehdolla konkordanssihaku tuottaa 30 osumaa, joista puolet on valideja. Esimerkissä 137 määritteenä on partisiippimuoto *myötävaikuttava*. Koettimen jälkeen su- luissa luetellut vaikuttavat tekijät vaikuttavat tapahtumaan onnettomuus onnettomuuden synnyn kautta.

- (137) Ideointivaiheessa tunnistettiin varsin suuri joukko erilaisia onnettomuusmahdollisuuksia ja onnettomuuden syntyyn myötävaikuttavia tekijöitä (pohjakosketus tai karilleajo, yhteentörmäys, kaatuminen tai vuoto tai uppoaminen, tulipalo tai räjähdys, vahingollisen aineen päästö). (smt99: 42)

Muut huomioon otettavat tekijän tai osatekijän määritteet ovat käsinoiminnan perusteella adjektiivit *tärkeä*, *keskeinen* ja *olennainen*. Hakuehto *tärkeä\* tekijä\*/tärkeä\* osatekijä\*/keskei\* tekijä\*/olennai\* tekijä\*/olennai\* osatekijä\** tuottaa yhdeksän osumaa, joista kahdeksan on valideja:

- (138) Myös korroosio- ja pinnoitekysymykset, tankkerikaluston kunnonvalvonta sekä tankkerikaluston ikääntyminen, ikärajat ja elinkaarisuunnittelu ovat tärkeitä tekijöitä riskien hallinnassa. (smr02: 59)

Esimerkissä 138 koettimena toimii adjektiivin *tärkeä* ja substantiivin *tekijä* yhdistelmä. Esimerkissä on lueteltu peräti kuusi toimintaan riskien hallinta vaikuttavaa tekijää: korroosio- kysymykset, pinnoitekysymykset, tankkerikaluston kulunvalvonta, tankkerikaluston ikääntyminen, tankkerikaluston ikärajat ja tankkerikaluston elinkaarisuunnittelu.

- (139) Jopa tankin ilmaputken kautta tapahtuva vuoto on ollut tärkeänä osatekijänä eräissä have- reissa. (smt99: 18)

Esimerkissä 139 koettimena on sanaliitto *tärkeä osatekijä*. Tässä esimerkissä vaikuttava tekijä on vuoto, joka mahdollisesti vaikuttaa haveriin. Tarkentamalla hakuehtoa määritellään osumalistasta karsiutuvat osumat, joissa tekijän määriteosana on adjektiivi *inhimillinen*. Merenkulun turvallisuuskontekstissa *inhimillinen tekijä* on termi, jolla on koetinkorpuksessa 23 esiintymää:

- (140) Kansainvälisesti on todettu, että n. 80 % onnettomuuksista on inhimillisen tekijän aiheuttamia ja vain n. 20 % teknisen vian aiheuttamia. (hv92: 12)

Substantiivi *tekijä* on hyvä vaikutussuhteen koetin määriteosalla tarkennettuna, jolloin roskan määrä pienenee oleellisesti. Koetinedokkaalla *tekijä* tehtyjen konkordanssihakujen tuottamien osumalistojen tarkastelu paljasti, että tämän koettimen yhteydessä vaikutus on usein vähentävä tai parantava (lisäävä tai heikentävä). Tämä lisää koettimen käyttökelpoisuutta, sillä vaikutuksen suuntaa ei tarvitse hakea kaukaa tekstikontekstista.

**Koetinedokas vaikutus.** Koettimien käsinpoiminnan perusteella substantiivi *vaikutus* on vaikuttavan tekijän mahdollinen substantiivikoetin, joka toimii koettimena hyvin samalla tavalla kuin verbikoetin *vaikuttaa*. Konkordanssihaku hakuehdolla *vaikutus/vaikutuksia* tuotti 44 osumaa, joista 18 on valideja.

- (141) Luotsauksen ohella on saattohinauksella merkittävä vaikutus turvallisuutta parantavana tekijänä. (smr02: 61)

Esimerkissä 141 vaikuttava tekijä on toiminta saattohinaus ja vaikutuksen kohde on ominaisuus turvallisuus, johon saattohinaus vaikuttaa positiivisesti. Esimerkissä mainitaan toinenkin turvallisuuteen positiivisesti vaikuttava tekijä: toiminta luotsaus.

- (142) Aluksen toimintakulttuurilla on suora vaikutus komentosiltatyön onnistumiseen. (kiv97: 92)

Esimerkissä 142 vaikuttava tekijä on aluksen toimintakulttuuri ja vaikutuksen kohde toiminta komentosiltatyö. Samoin kuin koetinedokkaan *vaikuttaa* tuottamissa valideissa tapauksissa myös substantiivin *vaikutus* tuottamissa tapauksissa tekijä vaikuttaa kohteeseensa usein epäsuorasti, kuten esimerkissä 142, jossa aluksen toimintakulttuuri vaikuttaa komentosiltatyön onnistumiseen eikä suoraan komentosiltatyöhön.

Hieman yli puolet hakuehdon *vaikutus/vaikutuksia* tuottamista osumista on roskaa. Roskan määrää saadaan vähennettyä, kun terminologista koetinta käytetään haussa yhdessä termin kanssa, sillä sekä vaikuttavan tekijän nimitys että vaikutuksen kohteen nimitys ovat hyvin lähellä koetinta. Vaikuttavan käsitteen nimitys on useimmiten välittömästi ennen koetinta ja vaikutuksen kohteen nimitys yleensä heti koetinedokkaan jälkeen. Vaikutuksen suunta on tekstissä kaikissa valideissa osumissa vasemmalta oikealle, eli vaikuttava käsite on ennen vaikutuksen kohdetta.

**Koetinedokkaat osa ja osuus.** Käsinpoiminnassa substantiivit *osa* ja *osuu*s osoittautuivat mahdollisiksi vaikutussuhteen koettimiksi. Konkordanssihaku hakuehdolla *osuu*s/*osa*\* tuotti kuitenkin peräti 398 osumaa, joten hakuehto on aivan liian laaja. Osumalistan lähempi tarkastelu osoitti, että substantiivien *osuu*s ja *osa* edessä on lähes aina määrite *tärkeä* tai *merkittävä*, silloin kun substantiivit toimivat vaikutussuhteen koettimina:

- (143) Onnettomuuksien ehkäisemisessä ja vahinkojen minimoinnissa esittää tärkeää osaa inhimillinen tekijä. (hv98: 3)

Esimerkissä 143 vaikuttava tekijä on inhimillinen tekijä, joka vaikuttaa toimintoihin onnettomuuksien ehkäiseminen ja vahinkojen minimoiminen. Hakuehto *tärkeä\* osa\** tuotti kahdeksan osumaa, joista kolme on valideja vaikutussuhteen tapauksia. Validien osumien määrä on niin pieni, ettei koetinehdokas ole kelvollinen vaikuttavan tekijän koetin ainakaan tämän tutkimuksen aineistossa. Hakuehdot *tärkeä osuus* ja *merkittävä osuus* tuottivat kumpikin vain yhden osuman, joten ne eivät ole kelvollisia vaikutussuhteen koettimia tutkimukseni aineistossa.

Osumalistasta nousi esiin sana *osaltaan*, jota kirjoittajat näyttävät käyttävän silloin, kun he puhuvat kahden käsitteen välisestä vaikutussuhteesta. Koetin *osaltaan* tuotti 15 osumaa, joista kuusi on validia:

- (144) VTS:n tehtävänä on lisätä liikenteen turvallisuutta ja vähentää riskejä. Siten se osaltaan auttaa myös estämään ympäristövahinkoja. (smr02: 65)

Validien osumien osuus kaikista osumista on kuitenkin alle puolet, joten suuremmasta aineistosta koetinehdokas tuottaisi todennäköisesti paljon roskaa. Sen määrää ei voi vähentää tunnetun termin ja koettimen yhdistelmähaulla, sillä valideissakaan osumissa ei koetimeen aina välittömästi liity termiä.

**Yhteenveto.** Suomalaisista vaikuttavan tekijän koetinehdokkaista verbi *vaikuttaa* osoittautui konkordanssihaussa produktiiviseksi, mutta se tuotti myös koko joukon kelpaamattomia osumia. Sen sijaan verbi *riippua* ei osoittautunut kovin produktiiviseksi tutkimukseni aineistossa. Substantiivikoettimista produktiivisin oli *tekijä*. Se tuotti kuitenkin runsaasti kelpaamattomia osumia. Niiden osuutta on mahdollista vähentää tarkennetulla haulla, mutta hakuehto muodostuu tällöin helposti erittäin pitkäksi. Substantiivi *vaikutus* osoittautui lähes yhtä produktiiviseksi kuin *tekijä*, mutta sen hakuehto on huomattavasti lyhyempi *vaikutus/vaikutuksia*. Substantiivin *osa* muodoista produktiivisimmaksi osoittautui *osaltaan*. Konkordanssihaku mahdollisilla koettimilla tuotti vaikuttavan tekijän koettimien listaan verbit *parantaa* ja *vähentää*, jotka osoittautuivat tutkimukseni aineistossa produktiivisiksi ja kohtuullisesti roskaa tuottaviksi vaikuttavan tekijän koettimiksi.

#### 7.4.7.3 Toimintasuhteiden suomalaiset koettimet

Nuopposen (2008: 26) uudistetussa käsitesuhteiden luokittelussa toimintasuhteet (activity relations) kuuluvat funktionaaliin käsitesuhteisiin (functional concept relations). Nuopponen jaottelee toimintasuhteet edelleen muun muassa toiminnan ja toimijan väliseen agenttiivisuhteeseen (agent relation), toiminnan ja toiminnan kohteen väliseen objektisuhteeseen (object relation) sekä toiminnan ja toiminnan välineen väliseen välinesuhteeseen (tool relation). Toimintasuhteiden rinnalla Nuopposen jaottelussa ovat instrumentaalisuhteet (instrumental relations). Olen tässä tutkimuksessani kuitenkin luokitellut menetelmä- ja keinosuhteet toimintasuhteisiin, niin kuin Nuopponenkin (2005: 135) on tehnyt aiemmassa käsitesuhdeluokittelussaan. Koetinhaku aineistosta *pw02* ja koettimien käsinpoiminta aineistosta *fi12* tuotti seitsemän tämän jaottelun mukaisten toimintasuhteiden koetinehdokasta. Jotkut koetinehdokkaat toimivat kahden käsitesuhteen koettimena. Esimerkiksi *avulla* on sekä välineen että keinon koetin. Analyysissani välinesuhde eroaa keinosuhteesta siten, että välinesuhteen väline on olio kun

taas keinosuhteen keino on toiminta. Tutkimukseni tuottamat toimintasuhteen terminologiset koettimet olen koonnut taulukkoon 47.

TAULUKKO 47. Toimintasuhteen terminologiset koettimet suomenkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
käyttää	<i>käyttäen</i>	13	13	100	vahva
menetelmä	<i>menetelm*</i>	72	16	22	heikko
keino	<i>keino*</i>	26	19	73	vahva
apu	<i>avulla</i>	120	63	53	vahva
suorittaa	<i>suorittaa/suorittavat/suorittama/suorittamia</i>	29	7	24	heikko
osallistua	<i>osallistuu</i>	8	6	75	hyvä
kohdistua	<i>kohdistuu/kohdistua/kohdistuva*</i>	34	13	38	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Menetelmäsuhde.** Manuaalinen koetinhaku aineistosta *fi12* tuotti kaksi mahdollista menetelmän koetinta, jotka ovat verbin *käyttää* 2. infinitiivin instrumentaalimuoto *käyttäen* ja substantiivi *menetelmä*. Näiden koetinehdokkaiden testaamiseksi tein konkordanssihaut aineistosta *fi12* hakuehdoilla *käyttäen* ja *menetelm\**.

**Koetinehdokas *käyttää*.** Konkordanssihaku aineistosta *fi12* mahdollisella menetelmän verbikoettimella *käyttäen* tuotti 13 osumaa, jotka kaikki ovat valideja menetelmän tai välineen tapauksia. Tosin yhdessä tapauksessa menetelmän kohteena olevan toimintakäsitteen nimitys on konkordanssihaun tulostusalueen ulkopuolella. Kuudessa validissa tapauksessa menetelmä tai väline on välittömästi ennen koetinehdokasta:

(145) Esitys on vain suuntaa-antava ja tarkemmin syyseurausketjun voi hahmottaa vaikka vika-puumenettelyä käyttäen. (smr02: 46)

Lähes yhtä yleinen on tapaus, jossa välineen tai menetelmän nimitys on välittömästi koetinehdokkaan jälkeen, kuten esimerkissä 146 välineen nimitys *automaattiohjaus*:

(146) Toinen vireystilaan liittyvä asia on ajo yksin komentosillalla avomerellä käyttäen automaattiohjausta. (kiv97: 78)

**Koetinehdokas *menetelmä*.** Menetelmän mahdollinen substantiivikoetin *menetelmä* tuotti hakuehdolla *menetelm\** peräti 72 osumaa aineistosta *fi12*. Substantiivi *menetelmä* ei kuitenkaan ole vahva koetin, sillä vain 16 osumassa koetinehdokas toimii koettimena ja 56 osumassa se on muussa tehtävässä, sillä *menetelmä* on yleinen sana, jonka merkitys riippuu kontekstista. Tekstien kirjoittajat saattavat todeta yleisesti, että monta menetelmää on käytössä, tai vaativat niiden kehittämistä, mutta menetelmiä ei kuvata likikään aina. Sanaa *menetelmä* käytetään myös hyvin usein väliotsikoissa, mikä kasvattaa roskaosumien määrää. Lisäksi sanaa käytetään yhteyksissä, joissa sen tilalle sopisi paremmin esimerkiksi sana *toimintatapa*, kuten esimerkissä 147:

- (147) Kuten aiemmin on viitattu VTS säännöksiä ja menetelmiä kehitettiin lähes pelkästään satamakohtaisesti satamaviranomaisten toimesta, eikä edes kansallisia säännöksiä ollut. (pw02: 2)

Valideista osumista seitsemässä kontekstissa on mainittu sekä menetelmä että toiminta, johon menetelmää käytetään:

- (148) Suomen olosuhteisiin soveltuisi ns. Tuck & Taylorin menetelmä matalikon aiheuttaman imun (squat) laskemiseksi (Norrbin 1983). (kiv97: 82)

Esimerkissä 148 menetelmän nimitys on *Tuck & Taylorin menetelmä*, joka on välittömästi ennen koetinehdokasta *menetelmä*. Itse asiassa koetinsana on osa menetelmän nimitystä, kuten 11 muussakin validissa tapauksessa. Esimerkin 148 mukaan menetelmän kohde on toiminta matalikon aiheuttaman imun laskeminen. Joissakin tapauksissa toimintakäsitteen nimitys jää konkordanssihaun tulostusalueen ulkopuolella:

- (149) Tässä kuvattu tapaustutkimus tarjosi mahdollisuuden kokeilla aiemmin teollisuuden kohteissa menestyksekkäästi käytettyjä järjestelmällisen turvallisuusanalyysin menetelmiä nimenomaisesti laivaympäristöön sovellettuna. (smt99: 38)

Esimerkissä 149 kirjoittaja mainitsee järjestelmällisen turvallisuusanalyysin menetelmät, mutta toimintaa, johon niitä sovelletaan, ei virkkeessä mainita.

**Välinesuhde ja keinosuhde.** Koettimien käsinpoiminnassa aineistosta *fil2* mahdolliseksi välinesuhteen tai keinosuhteen koettimeksi osoittautui substantiivin *apu* yksikön adessiivimuoto *avulla*. Konkordanssihaku hakuehdolla *avulla* tuotti 120 osumaa. Osumalistan manuaalisen tarkastelun perusteella 63 osumassa tarkoitteeseen vaikutetaan jollakin välineellä tai keinolla:

- (150) Järjestelmän avulla voidaan liikennetilanteeseen vaikuttaa niin, että liikenne on mahdollisimman turvallista ja sujuvaa. (kiv97: 85)

Esimerkissä 150 vaikuttamisen väline on järjestelmä ja vaikuttamisen kohde liikennetilanne. Esimerkissä 151 välinekäsite on *VTS-keskus*, joka on välinesuhteessa toimintakäsitteeseen *alusliikenteen ohjaus*, joka on esimerkissä ilmaistu verbitermillä *ohjata alusliikennettä*:

- (151) Kansallisille merialueille ehdotetaan perustettavan VTS-keskuksia, joiden avulla kyetään ohjaamaan alusliikennettä turvallisesti. (smr02: 97)

Väline- tai keinosuhteen koetin *avulla* on tekstissä lähes aina välittömästi välinettä tai keinoa nimeävän termin jälkeen:

- (152) Suomen ympäristökeskus pyrkii estämään alusten tahallisia öljypäästöjä valvonnan avulla. (smr02: 71)

Esimerkissä 152 keinona on toiminta, jonka nimitys *valvonta* on esimerkkikontekstissa välittömästi ennen koetinehdokasta. Kymmenessä tapauksessa termin ja koetinehdokkaan välissä on kuitenkin pilkku ja *joka*-pronomini, kuten esimerkissä 155. Toimintakäsitteen nimitys on joko ennen koetinehdokasta, kuten esimerkissä 156 *alusten tahallisten öljypäästöjen estäminen*, tai koetinehdokkaan jälkeen kuten esimerkissä 155 *ohjata*

*alusliikennettä*, mutta toimintakäsitteen nimityksen ja koetinehdokkaan välissä on yleensä nimitykseen kuulumattomia sanoja.

**Koetinehdokas *keino*.** Substantiivi *keino* osoittautui käsinpoiminnassa mahdolliseksi keinosuhteen koettimeksi. Konkordanssihaku aineistosta *fi12* hakuehdolla *keino\** tuotti 26 osumaa, joista 19 on validia tapausta. Esimerkissä 153 *keino* on toiminta AIS-järjestelmän käyttöönotto, jolla vaikutetaan turvallisuuteen. Toivottu vaikutus on turvallisuuden parantuminen:

(153) Myös AIS-järjestelmän käyttöönotto olisi yksi keino parantaa turvallisuutta. (sms 02: 35)

11 validissa tapauksessa on mainittu ainakin yksi *keino*, kuten esimerkissä 153. Viides-ssä validissa tapauksessa käsitteiden nimitykset ovat tekstissä samassa järjestyksessä kuin esimerkissä 153, eli keino nimitys on ennen koetinehdokasta ja sen avulla toteutettavan toiminnan nimitys välittömästi koetinehdokkaan jälkeen. Järjestys voi olla myös päinvastainen. Seitsemässä validissa tapauksessa ei keinoja ole paljastettu ainakaan osumalistaan tulostuneessa kontekstissa, mutta sanamuodosta käy ilmi, että keinoja on olemassa:

(154) Tässä ei kuitenkaan tehdä riskianalyysiä, tarkastellaan vain hieman keinoja vähentää ympäristöriskejä. (smr02: 75)

Yhdessä tapauksessa on epävarmaa, löytyykö yhtään keinoja. Kahdessa validissa tapauksessa kirjoittaja sanoo, ettei keinoja ole tai ne ovat heikot, mikä sekin saattaa olla tarpeellinen tieto:

(155) Esimerkiksi aallon korkeuden tai jään tarkan laadun mittaamiseen ei ole olemassa täsmällisiä keinoja. (kiv97: 96)

Esimerkissä 155 kirjoittaja kertoo, että toimintoihin aallon korkeuden mittaaminen ja jään laadun mittaaminen ei ole ainakaan täsmällisiä keinoja. Aineistoa kuvaa havainto, että viidessä validissa tapauksessa toiminta, johon etsitään keinoja, on turvallisuuden parantaminen.

**Agenttiivisuhde.** Toiminnan määritelmästä johtuu, että toiminnalla on aina oltava agentti. Toimintakäsitteen agentti voi olla yksilö, yksilöistä koostuva ryhmä tai kollektiivinen toimija. Kollektiivisia agenteja voivat olla esimerkiksi hallitus, viranomainen tai organisaatio. Jotta tällainen kollektiivinen agentti voisi toimia, on tiettyjen sosiaalisten ja normatiivisten ehtojen täytyttävä. (Pilke 2000: 115.)

Aina ei toimintakäsitteen agenttia kuitenkaan ilmoiteta edes sanakirjamääritelmissä. Pilke (2000: 190) havaitsi yllätyksekseen, että hänen tutkimusaineistossaan tekniikan toimintakäsitteiden sanakirjamääritelmistä agentti puuttuu kokonaan. Tekniikan alalla tavallisia toimintakäsitteitä ovat prosessit, joissa ei Pilkkeen tulkinnan mukaan ole tarpeen tietää, kuka toiminnan suorittaa. Sen sijaan prosessi itsessään on tärkeä. Yleinen ei agentti ole Pilkkeen mukaan lääketieteen tai juridiikan sanakirjamääritelmissäkään. Lääketieteen toimintakäsitteiden määritelmissä agentti oli mainittu viidessä ja juridiikan käsitteiden määritelmissä 18 prosentissa (Pilke 2000: 191). Oletettavaa on, että tekstiaineistossa agentti on mainittu toimintakäsitteen yhteydessä tätäkin harvemmin.

Tämän tutkimuksen suomenkielisessä aineistossa on käsinpoiminnan perusteella 14 tapausta toiminnan ja agentin välisestä käsitesuhteesta. Aina ei ole selvää, onko käsitteiden välillä käsitesuhde vai onko toimija paremminkin toimintakäsitteen piirre. En-



nakko-oletuksen mukaisesti tapauksia on vähän, mutta tästä huolimatta on hämmästyttävää, että niitä on vain 14. Mahdollinen selitys on, että aineistossa on suhteellisen vähän toimintakäsitteitä. Aineistosta *fi12* tehdyn käsinpoiminnan perusteella mahdollisia agentin koettimia ovat verbit *suorittaa* ja *osallistua*. Konkordanssihaun tein hakuehdoilla *suorittaa/suorittavat/suorittama/suorittamia* ja *osallistuu*.

**Koetinehdokas *suorittaa*.** Hakuehto *suorittaa/suorittavat/suorittama/suorittamia* tuotti 29 osumaa, joista seitsemän on valideja:

(156) Katsastuksia suorittavat sivutoimiset palkkiopohjalla toimivat katsastajat. (hv92: 14)

Esimerkissä 156 koetinehdokas *suorittavat* osoittaa toimintakäsitteen *katsastus* ja agentin *katsastaja* välistä suhdetta. Hakuehdon osa *suorittaa* tuotti runsaasti lauseita, joissa tekijää ei mainita:

(157) Esimerkiksi GOFREP-järjestelmän vaatiman ennakoilmoittautumisen voi suorittaa AIS-laitteen lähettämien perusviestien avulla. (mv10: 12)

Esimerkissä 157 koetinehdokas *suorittaa* on osa toimintakäsitettä *suorittaa ennakoilmoittautuminen*, mutta tekijää eli suorittajaa ei mainita. Viidessä osumassa koetinehdokas *suorittaa* liittyy substantiiviin *tutkinto*:

(158) Päälikkö saattoi uuden asetuksen mukaan itse suorittaa linjaluotsin tutkinnon. (kiv97: 70)

Agentti on ennen koetinehdokasta ja toimintakäsite sen jälkeen kolmessa tapauksessa, ja neljässä tapauksessa järjestys on päinvastainen. Kaikissa valideissa tapauksissa on jommankumman käsitteen nimitys välittömästi ennen koetinehdokasta, ja neljässä tapauksessa kummankin käsitteen nimitys liittyy välittömästi koetinehdokkaaseen.

**Koetinehdokas *osallistua*.** Hakuehto *osallistuu* tuotti vain kahdeksan osumaa mutta niistä kuusi on validia:

(159) Rajavartiolaitos osallistuu aktiivisesti poliisin johtamaan, merialueella tapahtuvaan, terrorismin torjuntaan. Tulli osallistuu yhteistyöhön etenkin tavara- ja konttiliikenteen valvonnan osalta. (sms02: 31)

Esimerkissä 159 agentti rajavartiolaitos on kollektiivinen toimija, joka on agenttina toiminnalle terrorismin torjunta yhdessä kahden muun kollektiivisen agentin, poliisin ja tullin, kanssa. Agentin nimitys on välittömästi ennen koetinehdokasta kaikissa valideissa tapauksissa yhtä lukuun ottamatta.

Kummankin hakuehdon tuottamissa valideissa tapauksissa agenteina on sekä yksilöitä (tarkastuspäälliköt, merenkululaitoksen tarkastajat, ruorimies, konemestari, katsastajat, merenkulkupiirin tarkastajat) että kollektiivisia toimijoita (laivapäällystö, komentositahenkilökunta, yhteiskunta, valtiot, merenkululaitos, luokituslaitos, tulli, rajavartiolaitos).

**Objektisuhteen koetinehdokas *kohdistua*.** Koetinehdokkaiden käsinpoiminta tuotti vain yhden mahdollisen objektisuhdetta ilmaisevan terminologisen koettimen, verbin *kohdistua* (*kohdistuu*, *kohdistuva*). Konkordanssihaku aineistosta *fi12* hakuehdolla *koh-*

*distuu/kohdistua/kohdistuva*\* tuotti 34 osumaa, joista 18 tapauksessa on mainittu toiminnan kohde:

- (160) Vika- ja vaikutusanalyysi (Failure Mode and Effects Analysis, FMEA), joka kohdistuu lähinnä teknisen järjestelmän vikoihin - - (smr02: 50)
- (161) Henkilökuntaan eli siihen osaan, joka aiheuttaa noin 80 % onnettomuuksista kohdistuu 15 % valvonnasta. (hv92: 14)

Esimerkissä 160 teknisen järjestelmän vika on toiminnan vika- ja vaikutusanalyysi kohteena. Kohde on tekstissä lähes yhtä usein koetinedokkaan jälkeen, kuten esimerkissä 160, kuin ennen koetinedokasta. Esimerkissä 161 toiminnan kohde, henkilökunta, on ennen koetinedokasta ja toimintakäsitteen nimitys *valvonta* välittömästi koetinedokkaan jälkeen. 13 tapauksessa ainakin toisen käsitteen nimitys liittyy välittömästi koetinedokkaaseen.

Konkordanssihaku tuotti kahdeksan osumaa, joissa toiminnan tilalla on tapahtuman mahdollisuus, joka on ilmaistu sanoilla *vaara*, *riski* tai *uhka*:

- (162) Vaarat voivat kohdistua ihmisiin, ympäristöön tai omaisuuteen. (smr02: 50)

Esimerkissä 162 oliot ihminen, ympäristö ja omaisuus ovat vaaran kohteena.

**Yhteenveto.** Menetelmän koetinedokkaista substantiivi *menetelmä* osoittautui produktiiviseksi mutta roskaa tuottavaksi ja verbi *käyttäen* sekä produktiiviseksi että vähän roskaa tuottavaksi. Väline- tai keinosuhteen koetinedokas *avulla* osoittautui erittäin produktiiviseksi ja kohtalaisen tarkaksikin koettimeksi. Toinen keinosuhteen koetinedokas, substantiivi *keino*, ei ollut yhtä produktiivinen mutta sen tarkkuus oli hyvä. Kahdesta agenttiivisuhteen koetinedokkaasta kumpikaan ei osoittautunut vahvaksi koettimeksi. Verbit *suorittaa* ja *osallistua* tuottivat liki yhtä monta validia osumaa, mutta verbi *suorittaa* tuotti huomattavasti enemmän roskaa kuin verbi *osallistua*, joka tuotti vain kaksi roskaosumaa. Tästä syystä *osallistua* on vahvin toimijan koetin tutkimukseni aineistossa. Verbin *suorittaa* käyttökelpoisuus kuitenkin paranee, jos haussa on mahdollista käyttää tunnettua termiä, sillä käsitteen nimitys on aina välittömästi ennen koetinedokasta.

Tutkimukseni aineistossa on vain vähän konteksteja, joissa on toiminnan ja toiminnan kohteen välinen käsitesuhde. Koetinedokkaan *kohdistua* tuottamista osumista noin kolmannes on valideja toiminnan ja toiminnan kohteen välisen suhteen konteksteja. Koska osumissa on enemmän roskaa kuin valideja osumia, ei koetinedokas ole vahva koetin. Konkordanssihaku paljasti myös, että koetinedokas tuottaa tapauksia, joissa ei ole toimintasuhdetta vaan tapahtuman mahdollisuuden ja olion välinen käsitesuhde.

## 7.5 Venäläiset terminologiset koettimet

Aloitin venäläisten terminologisten koettimien tunnistamisen samalla tavalla kuin suomalaisten koettimien tunnistamisen, jota olen selostanut edellä jaksossa 7.4. Koettimien poimintakokeen ensimmäisenä venäjänkielisenä aineistona oli 2 793 sanan artikkeli *Eskos*, josta venäläiset asiantuntijat olivat poimineet termit. Tein ensin tekstistä konkordanssihaun tunnetuilla termeillä ja kokosin alustavan venäläisten koetinedokkaiden luettelon. Konkordanssihaku tuotti 109 sellaista kontekstia, jossa jokin koetinedokas ilmaisee terminologisesti kiinnostavaa tietoa. Seuraavaksi tuotin konkordanssilistoista

poimittujen koetinehdokkaiden konkordanssit ja tutkin, kuinka usein koetinehdokas todella toimii koettimena ja kuinka usein jossakin muussa tehtävässä.

*Eskos*-artikkelista tehdyn haun jälkeen jatkoin koetinehdokkaiden tutkimusta testaamalla niitä laajemmalla venäläisellä aineistolla, josta käytän lyhennettä *Bezop*. Lisäksi täydensin koettimien luetteloa manuaalisella poiminnalla aineistosta *Bezop*. Manuaalinen poiminta tuotti 1 142 kontekstia, joissa jokin koetinehdokas ilmaisee terminologisesti kiinnostavaa tietoa. Manuaalinen poiminta tuotti myös uusia koetinehdokkaita, jotka ilmaisevat määritelmää, vaikuttavaa tekijää, toiminnan välinettä, menetelmää, tapaa tai kohdetta sekä tapahtuman syytä tai seurausta. Seuraavaksi tarkastelen venäjänkielisellä artikkelilla tekemäni konkordanssihaun ja laajemmalla aineistolla tekemäni manuaalisen poiminnan tuottamien koetinehdokkaiden toimintaa terminologisina koettimina ja niiden käyttökelpoisuutta terminologisesti kiinnostavan tiedon poiminnassa.

### 7.5.1 Käsitteen nimityksen venäläiset koettimet

Tässä jaksossa puhun termien sijasta käsitteen nimityksistä kuten tein suomalaisia käsitteen nimityksen koettimia selostaessani, sillä erikoisalan käsitteitä nimetään teksteissä termien ohella nimityksillä, jotka eivät esimerkiksi vakiintumattomuutensa takia ansaitse termin statusta, mutta ovat tästä huolimatta terminologisesti kiinnostavia (ks. jakso 7.4.1). Asiantuntijoiden poimimilla termeillä tehty konkordanssihaku venäjänkielisestä artikkelista *Eskos* ja konkordanssilistoihin tulostuneiden termikontekstien tarkastelu tuotti kaksi käsitteen nimitykseen viittaavaa venäläistä koetinehdokasta. Nämä ovat sulkeet ja substantiivi *концепция* (käsite). Haku ei tuottanut muita käsitteen nimityksen koetinehdokkaita, mutta englanninkielisillä aineistoilla tehtyjen tutkimusten perusteella myös substantiivit *понятие* (käsite) ja *термин* (termi) ovat todennäköisesti käsitteen nimityksen koettimia. Ne osoittautuivatkin hyviksi koetinehdokkaiksi aineistosta *Bezop* tehdyssä manuaalisessa koetinhausta. Substantiivikoettimien *понятие* ja *термин* lisäksi manuaalinen koettimien haku aineistosta *Bezop* tuotti mahdolliset käsitteen nimityksen verbikoettimet *называться* (olla nimeltään) ja *называть* (kutsua). Koetinehdokkaista sulkeet on poimintojen perusteella produktiivisin. Tästä syystä selostan ensin sulkeiden toimintaa käsitteen nimityksen koettimena tutkimukseni venäjänkielisessä aineistossa. Tämän jälkeen selostan mahdollisten substantiivikoettimien *понятие* ja *термин* sekä verbikoettimien *называться* ja *называть* toimintaa. Konkordanssihaku tulokset olen koonnut taulukkoon 48.

TAULUKKO 48. Käsitettä nimeävän ilmauksen terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
sulkeet	(*	370	134	36	hyvä
понятие	поняти*	33	20	61	vahva
термин	термин*	17	11	65	vahva
концепция	концепци*	14	8	57	vahva
называется	называется	10	7	70	hyvä
называемый	называем*	6	5	83	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas sulkeet.** Venäjänkielisestä artikkelista *Eskos* tehty konkordanssihaku hakuehdolla (\* tuotti 43 osumaa. Lähes joka toisessa eli 18 osumassa sulkeissa on vähintään yksi venäjänkielinen termi. Suurimmassa osassa valideista osumista eli 13 osumassa sulkeissa on termin lyhyt muoto. Viidessä osumassa on vähintään kaksi termiä ja lisäksi yhdessä osumassa sulkeissa on venäläisen termin englanninkielinen vastine. Artikkelilla *Eskos* tehdyn haun perusteella sulkeet ovat myös hyvä synonyymisen ilmauksen koetin, sillä kymmenessä osumassa termin pitkä muoto on tekstissä välittömästi ennen sulkeissa olevaa lyhennetermiä:

- (163) Предпосылками создания единой системы контроля и управления судоходством (ЕСКУС), определяющими ее правовую основу, являются международные и российские программы и проекты. (Eskos: 15)

Esimerkissä 163 sulkeissa on lyhennetermi *ЕСКУС*, jonka pitkä muoto *единая система контроля и управления судоходством* (alusliikenteen ohjaus- ja informaatiopalvelut) on välittömästi ennen sulkeita.

Hakuehdon (\* tuottamista 43 osumasta yhdeksässä on sulkeissa jostakin organisatiosta tai dokumentista käytetty isokirjainlyhenne:

- (164) На основе Резолюции Международной морской организации (ИМО) А.801(19) от 23.11.1995 г. в системе предусмотрено создание морских районов А1 и/или А2 на подходах к основным портам РФ, в том числе Тихоокеанского бассейна. (Eskos: 15)

Samoin yhdeksässä osumassa on sulkeissa maan, kaupungin, meren tai sataman nimi. Viidessä osumassa sulkeissa on numero tai täydentävä huomautus ja yhdessä osumassa määritelmä.

*Bezop*-aineistosta hakuehto (\* tuotti 327 osumaa. Noin kolmasosassa eli 115 osumassa sulkeissa on termin lyhyt muoto, suureen tunnus, venäläisen termin englanninkielinen vastine tai yksi tai useampi termi, kuten esimerkissä 165:

- (165) Огнетушащие вещества — огнегасительные средства пожарных систем (пенообразователи, порошки, хладоны, газы). (Bezop: 62)

Esimerkissä 165 sulkeissa on käsitteen *огнетушащее вещество* (sammute) neljän alakäsitteen nimitykset *пенообразователь* (vaahdote), *порошок* (jauhe), *хладон* (halon) ja *газ* (kaasu). Termin ohella sulkeiden sisällä saattaa olla sanoja tai lyhenteitä, jotka eivät kuulu termiin, kuten esimerkissä 166 ilmaus *в виде*:

- (166) Эти перекрытия могут быть вертикальными (в виде переборок) и горизонтальными (в виде палуб). (Bezop: 54)

Viidessä tapauksessa sulkeissa on suureen tunnus (ks. esimerkki 167) ja vain kahdessa tapauksessa englanninkielinen vastine, kuten esimerkissä 168.

- (167) В этом случае на остойчивость отрицательно и весьма существенно влияет свободная поверхность воды в трюме и повышение общего центра тяжести (Z<sub>g</sub>) в результате образования слоя воды над контейнерами (случай с отсеком второй категории). (Bezop: 111)

- (168) По требованию Конвенции SOLAS-74 на каждом судне должен быть план противопожарной защиты (fire plan). - - (Bezop: 82)

Aineistossa *Bezop* termin lyhyt muoto on sulkeissa välittömästi termin jälkeen vain viidessä tapauksessa. Määrä on huomattavan alhainen verrattuna suppeampaan aineistoon *Eskos*, jossa termin lyhyt muoto on sulkeissa 13 tapauksessa. Tämä johtunee tekstien erilaisesta luonteesta. Aineisto *Bezop* on oppikirja ja *Eskos* asiantuntijoiden toisille asiantuntijoille kirjoittama aikakauslehden artikkeli.

**Koetinehdokas *концепция*.** Aineistolla *Eskos* tehdyn haun perusteella substantiivi *концепция* (käsite) on sulkeiden ohella mahdollinen käsitteen nimityksen koetin. Konkordanssihaku aineistosta *Eskos* hakuehdolla *концепци*\* tuotti kuusi osumaa, joista kolmessa sana *концепция* toimii koettimena. Esimerkissä 169 koetinehdokas viittaa lyhennetermiin *VTMIS*, joka on tekstissä välittömästi koetinehdokkaan jälkeen:

- (169) Однако в концепции VTMIS явно выражен информационный аспект, который должен найти отражение в разрабатываемых концепциях ЕСКУС с целью обмена информацией между локальными или региональными элементами ЕСКУС в Тихоокеанском бассейне, предоставления информации государственным структурам, обеспечивающим контроль над судоходством, и другим участникам транспортного процесса. (Eskos: 16)

Substantiivi *концепция* voi kuitenkin olla tekstissä muussakin tehtävässä kuin käsitteen nimityksen koettimena. Esimerkissä 169 sanaa *концепция* on käytetty toistamiseen, mutta nyt pikemmin merkityksessä 'luonnos'.

Laajemmasta aineistosta *Bezop* tehty konkordanssihaku hakuehdolla *концепци*\* tuotti kahdeksan osumaa, joista viisi on validia:

- (170) Понимание в свою очередь возможно только при наличии целостной философской концепции безопасности. (Bezop: 28)

Osumien määrä on yllättävän pieni, sillä 2 793 sanan artikkelissa sanan esiintymiä on vain kaksi vähemmän kuin 38 232 sanan oppikirja-aineistossa.

**Koetinehdokas *понятие*.** Aineistolla *Bezop* tehty manuaalinen koetinhaku tuotti substantiivin *концепция* lisäksi kaksi muuta mahdollista käsitteen nimityksen substantiivikoetinta. Nämä ovat substantiivit *понятие* (käsite) ja *термин* (termi). Konkordanssihaku aineistosta *Bezop* hakuehdolla *поняту*\* tuotti 33 osumaa, joista 20 on validia. Koetinehdokkaaseen *поняту*\* liittyvän käsitteen nimitys on lainausmerkeissä peräti 16 tapauksessa eli 80 prosentissa valideista osumista. Esimerkissä 171 lainausmerkeissä on termi *авария* (haveri):

- (171) Уже одно это свидетельствует об отсутствии единого толкования понятия "авария" и о необходимости его унификации, -- (Bezop: 27)

Vieläkin tyypillisempää koetinehdokkaalle on, että se on välittömästi ennen käsitteen nimitystä. Näin on 17 validissa tapauksessa. Muissa kuin valideissa osumissa keskustellaan yleisesti käsitteiden määrittelyn tarpeellisuudesta, kuten esimerkiksi 172, jossa kirjoittaja peräänkuuluttaa käsitteiden ja termien konkretisoimista:

- (172) Представляется, что нормативная база должна формулировать, конкретизировать понятия и термины, принципы, устанавливать нормативы безопасности и критерии оценки, фиксировать запреты и все это закреплять юридически. (Bezop: 34)

**Koetinehdokas *термин*.** Toinen manuaalisen koetinhaun tuottama käsitteen nimityksen mahdollinen substantiivikoetin, substantiivi *термин* (termi), ei konkordanssihaussa osoittautunut aivan yhtä produktiiviseksi kuin substantiivi *понятие*, sillä konkordanssihaku *Bezop*-aineistosta hakuehdolla *термин*\* tuotti vain 17 osumaa. Näistä 65 prosenttia eli 11 osumaa on validia, joten *термин* kuuluu ilman muuta vahvojen koettimien joukkoon. Koetinehdokas *термин* käyttäytyy hyvin samalla tavalla kuin koetinehdokas *понятие*, sillä seitsemässä tapauksessa se on välittömästi termin edellä ja termi on lainausmerkeissä:

(173) - - 11% не понимают термина "безопасная скорость" - - (Bezop: 7)

Esimerkissä 173 koetinehdokas on välittömästi ennen lainausmerkeissä olevaa termiä *безопасная скорость* (turvallinen nopeus). Neljässä tapauksessa koetinehdokkaaseen liittyy pronomini *этот* (tämä). Näissä tapauksissa termi on tekstissä niin kaukana ennen koetinehdokasta, että se ei näy konkordanssiohjelman perusasetuksilla tuotetussa kontekstissa, jossa näkyy kuusi sanaa ennen hakulauseketta. Lukijan on palattava tekstikontekstiin ja konkordanssihaun tuottamaa kontekstia edeltävään virkkeeseen, jossa käsitteen nimitys on mainittu. Esimerkissä 174 etsitty nimitys on *человеческий фактор* (inhimillinen tekijä):

(174) В настоящее время широко используется понятие "человеческий фактор". Однако анализ показывает, что специалисты по-разному понимают этот термин. (Bezop: 40)

Koetinehdokas *термин* käyttäytyy siis samalla tavoin kuin koetin *понятие*, joka sekin on tekstissä useimmiten välittömästi termin edellä, termi on usein lainausmerkeissä ja siihen saattaa liittyä pronomini *этот* (tämä). *Термин* käyttäytyy samalla tavoin kuin *понятие* myös sikäli, että muissa kuin valideissa osumissa puhutaan termeistä yleisesti.

**Koetinehdokkaat *называться* ja *называть*.** Substantiivikoettimien *понятие* ja *термин* lisäksi manuaalinen koetinhaku tuotti kaksi mahdollista käsitteen nimityksen verbikoetinta. Nämä ovat verbin *называться* preesensin yksikön 3. persoonan muoto *называется* (on nimeltään) ja verbin *называть* partisiippimuoto *называемый* (kutsuttu). Konkordanssihaku *Bezop*-aineistosta hakuehdolla *называется* tuotti 10 osumaa, joista seitsemän on validia. Kaikissa tapauksissa termi on tekstissä joko välittömästi koetinehdokkaan *называется* jälkeen tai välittömästi sitä ennen. Useimmiten koetinehdokas on ennen termiä, kuten esimerkiksi 175, jossa se on ennen termin *таранная переборка* (törmäyslaipio) elliptistä muotoa *таранная*:

(175) Первая от форштевня переборка называется таранной, она, как правило, образует форпик. (Bezop: 47)

Kaikissa valideissa tapauksessa koetinehdokkaan kontekstissa on termin lisäksi myös käsitteen määritelmä, kuten esimerkiksi 175, jossa käsitteen *таранная переборка* määritelmä on välittömästi ennen koetinehdokasta. Verbimuoto *называется* toimiikin yhtä hyvin määritelmän kuin termin koettimena.

Verbin *называться* lisäksi verbin *называть* partisiippimuoto *называемый* (kutsuttu) on manuaalisen koetinhaun perusteella mahdollinen käsitteen nimityksen koetin. Konkordanssihaku *Bezop*-aineistosta hakuehdolla *называем\** tuotti kuusi osumaa, joista viisi on validia tapausta koetinehdokkaan toiminnasta termin koettimena:



(176) В этом случае скуловая часть корпуса и междудонный лист образуют углубления для стока воды, называемые льялами. (Bezop: 47)

(177) При попадании сжиженного газа в воду наблюдается явление так называемого "беспламенного взрыва". (Bezop: 104)

Esimerkissä 176 koetinehdokas on välittömästi ennen termiä *ляло* (pilssi). Kaikissa viidessä validissa tapauksessa termi onkin tekstissä välittömästi koetinehdokkaan jälkeen. Kolmessa tapauksessa koetinehdokasta edeltää termin nimeämän käsitteen määrittelmä, kuten esimerkissä 176, jossa käsite *ляло* on määritelty ensin ja käsitettä nimeävä termi on annettu koettimen jälkeen. Esimerkissä 177 koetinta edeltää olosuhdepiirre nestekaasun joutuminen veteen, jonka vallitessa tapahtuma liekitön räjähdys (беспламенный взрыв) on mahdollinen. Koettimeen liittyy sana *так* (niin), ja termiehdokas on lainausmerkeissä. Sekä ilmaus *так называемый* että lainausmerkit osoittavat, että ainakaan kirjoittajan mielestä sanaliitto *беспламенный взрыв* ei ole vakiintunut termi.

**Yhteenveto.** Tutkimukseni perusteella koetin *понятие* on produktiivisin käsitteen nimityksen venäläisistä substantiivikoettimista. Se tuotti kaksi kertaa niin paljon osumia kuin toiseksi produktiivisin substantiivikoetin *термин*. Kummankin koettimen tuottamista osumista noin 2/3 on valideja. Substantiivi *концепция* ei ollut yhtä produktiivinen mutta senkin tuottamista osumista vähintään puolet on valideja. Se saattaa kuitenkin olla kirjoittajakohtainen. Kaikki käsitteen nimityksen substantiivikoettimet ovat vahvoja koettimia. Hyviä ovat myös verbikoettimet *называться* ja *называть*, jotka eivät ole kovin produktiivisia mutta tuottavat vain vähän roskaa. Sen sijaan sulkeet on käsitteen nimityksen koettimena ongelmallinen, sillä se on erittäin produktiivinen mutta alle puolet osumista on valideja. Parhaiten ne toimivatkin yhdessä tunnetun lyhennetermin kanssa. Sulkeiden ja lyhennetermin yhdistelmällä voidaan teksteistä hakea synonyymisia termin avattuja muotoja.

## 7.5.2 Synonymian venäläiset koettimet

Tässä jaksossa käsittelen koettimia, jotka ilmaisevat synonymiaa eli samaa käsitettä nimeävien termien välistä suhdetta. Totesin suomalaisten synonymian koettimien käsittelyn yhteydessä jaksossa 7.4.2, että olen tutkimuksessani ymmärtänyt synonymian laajasti, niin että se sisältää yhden kielen termien suhteen lisäksi myös erikielisten termien vastinesuhteen. Tästä samasta näkökulmasta käsittelen myös venäläisiä synonymian koettimia, jotka olen koonnut taulukkoon 49.

TAULUKKO 49. Synonymian terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
sulkeet	(*)	99	21	21	heikko
т.е.	т.е.	19	1	5	heikko
ИНЫМИ СЛОВАМИ	иными словами	6	1	17	heikko

<sup>a</sup> Ks. jakso 7.3.10.

Englanninkielisillä aineistoilla tehdyissä tutkimuksissa on todettu, että sulkeet on vahva synonymian koetin (ks. esim. Ahmad & Rogers 2001; Pearson 1998). Kuten olen kertonut jaksossa 7.4.2, sulkeet osoittautuivat vahvaksi synonymian koettimeksi myös tämän tutkimuksen suomenkielisessä aineistossa. Venäjänkielisestä artikkelista *Eskos* hakuehdolla (\* tehty konkordanssihaku osoittaa, että sulkeet ovat synonymian koetinehdokas myös venäjänkielisessä aineistossa. Konkordanssihaku tuotti 43 osumaa, joista peräti joka toisessa sulkeissa on isokirjainlyhenne. Niistä yli puolet eli 13 on tapauksia, joissa sulkeissa on termin pitkän muodon sanoista muodostettu lyhennetermi. Nämä lyhennetermit ovat synonyymejä termin pitkälle muodolle. Tarkentamalla hakuehtoa muotoon (\*) roska määrä putoaa huomattavasti, sillä tarkennettu hakuehto tuottaa vain 22 osumaa, joiden joukossa ovat kaikki validit tapaukset. Kymmenessä tapauksessa kolmesta toista termin pitkä muoto on tekstissä välittömästi ennen sulkeissa olevaa lyhennettä:

- (178) В соответствии с Резолюцией ИМО А.857 (20) от 27.11.1997. г. и рекомендациями Международной ассоциации маячных служб Министерством транспорта РФ осуществляется программа создания и модернизации систем управления движением судов (СУДС) в основных портах Дальневосточного региона России. (*Eskos*: 15)

Esimerkissä 178 sulkeissa on lyhennetermi *СУДС* (VTS) ja termin pitkä muoto *система управления движением судов* (alusliikennepalvelu) on välittömästi ennen sulkeita. Muissa kuin valideissa osumissa sulkeissa on yleensä erisnimen lyhenne, paikka, organisaation nimi tai täydentävä selite.

*Eskos*-artikkelilla tehdyn konkordanssihaun perusteella sulkeet näyttäisivät toimivan synonymian koettimena myös venäjänkielisessä tekstissä, mutta *Bezop*-aineistossa lyhenne on sulkeissa termin pitkän muodon jälkeen vain viidessä tapauksessa, vaikka aineiston laajuus on sanoina mitattuna yli 13-kertainen artikkeliin *Eskos* verrattuna. Valideja tapauksia sulkeiden toiminnasta synonymian koettimena on liki 40 000 sanan *Bezop*-aineistossa kaikkiaan vain kahdeksan. Niistä kolmessa sulkeissa on synonyymien termi:

- (179) Учебный пластырь (пластырь Макарова) имеется на всех судах и служит для тренировки экипажа при отработке действий по борьбе с водой. (*Bezop*: 53)

Esimerkissä 179 sulkeissa on termi *пластырь Макарова* (Makarovin vuotomatto), joka on synonyymi välittömästi ennen sulkeita olevalle termille *учебный пластырь* (harjoitusvuotomatto).

Jaksossa 7.4.2 totesin, että suomalaisessa tutkimusaineistossani sulkeissa on usein englanninkielinen termi ja suomenkielinen termi on välittömästi ennen sulkeita. Sen sijaan venäläisessä tutkimusaineistossa englanninkielinen termi on sulkeissa vain kolmessa tapauksessa:

- (180) - - на каждом судне должен быть план противопожарной защиты (fire plan), - - (*Bezop*: 82)

Manuaalinen koetinhaku venäjänkielisestä tutkimusaineistosta ei tuottanut sulkeiden lisäksi muita synonymian koetinehdokkaita. Suomenkielisessä aineistossa synonymian koettimeksi osoittautui konjunktio *eli*, joten tein konkordanssihaut venäjänkielisestä aineistosta *eli*-konjunktion mahdollisilla vastineilla *то есть*, *т.е.*, *иными словами* ja *или*. Kuten odottaa saattaa, eniten osumia tuotti konjunktio *или*. Se ei kuitenkaan osoittautunut vahvaksi koettimeksi, sillä se on synonymian koettimen tehtävässä vain viidessä tapauksessa 232 osumasta. Toiseksi eniten osumia mahdollisista vastineista tuotti

lyhenne *m.e.*, mutta 19 osumasta vain yhdessä lyhenne toimii synonymian koettimena. Ilmaus *иными словами* tuotti kuusi osumaa, joista yksi on validi. Lyhenteen *m.e.* avattu muoto *то есть* ei tuottanut yhtään osumaa.

**Yhteenveto.** Tämän tutkimuksen aineistossa sulkeet näyttäisivät jäävän ainoaksi venäläiseksi synonymian koettimeksi. Alhaisen tarkkuuden vuoksi sekin on vain heikko koetin. Näin pienellä aineistolla tehtyjen konkordanssihakujen perusteella ei kuitenkaan voi sanoa, ettei muitakin venäläisiä synonymian koettimia olisi.

### 7.5.3 Määritelmän venäläiset koettimet

Tunnettujen termien kontekstin tarkastelu artikkelista *Eskos* osoitti, että sulkeet on mahdollinen määritelmän koetin. Manuaalinen haku *Bezop*-aineistosta tuotti mahdolliseksi määritelmän koettimiksi sulkeiden lisäksi verbimuodot *понимается* (ymmärretään; tarkoitetaan), *понимают* (tarkoitetaan), *представляет собой* (on), *называется* (kutsutaan; sanotaan) sekä ajatusviivan ja pronominin *это* (se) yhdistelmän. Näillä koetinhdokkailla tehtyjen konkordanssihakujen tulokset ovat taulukossa 50.

TAULUKKO 50. Määritelmän terminologiset koettimet venäjänkielisessä aineistossa

Koetinhdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
sulkeet	(*	370	29	8	heikko
понимать, по- ниматься	<i>понима*</i> ei: <i>понимать</i> , <i>что/понимани*</i> <i>понимается/понимают</i>	35 9	10 7	29 78	heikko hyvä
представлять собой	<i>представляет собой</i>	17	16	94	vahva
называться	<i>называется</i>	10	7	70	hyvä
ajatusviiva + это	– <i>это</i>	15	7	47	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinhdokas *sulkeet*.** Konkordanssihaku hakuehdolla (\*) tuotti venäjänkielisestä aineistosta 30 sellaista osumaa, jossa sulkeet toimivat määritelmätiedon koettimena. 23 tapauksessa sulkeissa on termi, jonka nimeämää käsitettä koskeva määritelmätieto on annettu ennen sulkeita:

(181) Огнегасящим веществом является смесь бромистого этила и хладона (тетрафтордибромэтана) с добавками. (Bezop: 61)

Vain kahdeksassa määritelmässä on selkeästi annettu myös sulkeissa olevan käsitteen yläkäsite, kuten esimerkin 181 määritelmässä, jossa on annettu yläkäsite *смесь* (seos). Muissa tapauksissa konteksti sisältää käsitepiirteen mutta ei selkeää lähintä yläkäsitettä:

(182) При перевозке под небольшим давлением при частичном охлаждении (смешанный способ) расчетное давление в танках принимается 10 кгс/см<sup>2</sup>, а температура - 55°C. (Bezop: 105)

Viidessä tapauksessa sulkeissa on annettu määritelmä käsitteelle, joka on mainittu välittömästi ennen sulkeita:

- (183) Здесь нужны другие технические решения, и они были найдены в форме спасательных модулей (сбрасываемые шлюпки ныряющего типа), надежность которых оценивается уже в 70% (0,7). (Bezop: 11)
- (184) Одним из компонентов АИС является радиомаяк DGPS (станция для передачи дифференциальных поправок для глобальной спутниковой навигационной системы GPS Навстар/ Глонасс), обеспечивающий необходимую точность определения координат и вектора скорости. (Eskos: 16)

Sulkeiden toimintaa määritelmän koettimena heikentää roskaosumien suuri osuus, sillä hakuehto (\* tuotti venäläisestä tutkimusaineistosta 370 osumaa. Hakua olisi pystyttävä tarkentamaan, jotta manuaalisen työn osuus jäisi kohtuulliseksi. Osumien määrää on mahdollista rajoittaa muuttamalla hakuehtoa niin, että osumalistaan tulostuvat vain ne osumat, joissa sulkeissa on termi, koska suurimmassa osassa valideja tapauksia sulkeissa on termi eikä määritelmä. Hakuehdot ovat tällöin (\*) ja (\* \*), sillä kolmessa tapauksessa neljästä sulkeissa oleva termi on yksi- tai kaksisanainen.

**Koetinehdokkaat *понимать* ja *пониматься*.** Manuaalisen haun tulosten perusteella muotoiltu hakuehto *понима\** tuotti 35 osumaa aineistosta *Bezop*. Näistä 10:ssä koetinehdokas toimii määritelmän koettimena. Kahdeksassa validissa osumassa verbiin *понимать* (ymmärtää; käsittää) liittyy prepositio *под*. Suurin osa valideista osumista saadaan siis hakutulokseen tarkentamalla hakuehtoa preposition *под* ja verbimuotojen *понимается* tai *понимают* yhdistelmäksi. WordSmith 4.0 -konkordanssiohjelmassa tarkennettu haku ei kuitenkaan ole ongelmaton, sillä hakuehtoa muotoiltaessa pitäisi tietää, montako sanaa preposition ja verbin välissä on, koska jokainen näistä sanoista merkitään hakuehdossa katkaisumerkillä \*, jonka molemmilla puolilla on välilyönti. Tutkimukseni aineistossa verbin ja preposition välissä olevien sanojen määrä on 1–4 sanaa:

- (185) Поэтому под аварией традиционно понимают непредусмотренные расходы или убытки от повреждения или сами технические повреждения. (Bezop: 26)
- (186) Под конструктивной противопожарной защитой понимается комплекс пассивных средств, направленных на предотвращение возникновения пожаров; ограничение возникновения пожаров; ограничение распространения огня и дыма по судну; создание условий безопасной эвакуации людей из судовых помещений и с судна, а также для успешного тушения пожара. (Bezop: 54)

Esimerkissä 185 preposition ja koetinehdokkaan välissä on kaksi sanaa, joista välittömästi preposition jälkeen oleva on määriteltävän käsitteen nimitys. Esimerkissä 186 preposition ja koetinehdokkaan välissä on määriteltävän käsitteen nimitys, kuten suurimmassa osassa valideista tapauksista.

Hakuehdon *понима\** tuottamista osumista 11:ssä on ilmaus *следует понимать, что* (tulee ymmärtää, että). Yksikään näistä osumista ei ole validi tapaus koettimen toiminnasta määritelmän koettimena. Seitsemässä osumassa hakuehdon täyttää substantiivi *понимание* (ymmärtäminen; käsitys) eri taivutusmuodoissaan. Näissäkään osumissa ei ole käsitteen määritelmää, joten hakua voidaan tarkentaa siten, että sekä ilmaus *понимать, что* että substantiivi *понимани\** rajataan haun ulkopuolelle. Tarkkuutta voi

huomattavasti parantaa käyttämällä hakuehtoa *понимается/понимают*, mutta samalla menetetään noin kolmannes valideista osumista.

**Koetinehdokas *представлять*.** Konkordanssihaku hakuehdolla *представляет собой* (on) aineistosta *Bezop* tuotti 17 osumaa, joista 16 on valideja tapauksia koetinehdokkaan toiminnasta määritelmän koettimena. Määriteltävän käsitteen nimitys on joko välittömästi ennen koetinehdokasta, kuten esimerkissä 187, tai välittömästi koetinehdokkaan jälkeen, kuten esimerkissä 188:

(187) Пенопорошок представляет собой смесь сульфата алюминия  $Al_2(SO_4)_3$  с примесью каолина и бикарбоната натрия  $NaHCO_3$ , обработанного экстрактом лакричного корня. (Bezop: 59)

(188) Особый вид контроля за обеспечением безопасности мореплавания представляет собой судовой контроль. (Bezop: 35)

Määritelmät eivät kuitenkaan aina ole täydellisiä niin, että niissä olisi sekä yläkäsite että erottava piirre, vaan niissä saattaa olla pelkästään yläkäsite:

(189) Управление судном представляет собой информационный процесс, поэтому важное значение имеет количество информации, - - (Bezop: 39)

Koetinehdokas on kuitenkin hyvä, sillä lähes aina sen yhteydessä annetaan tietoa käsitteestä.

**Koetinehdokas *называться*.** Venäläistä koetinehdokasta *называется* (kutsutaan; sanotaan) vastaava englannin koetin *called* on yleensä koettimia tutkineiden aineistossa osoittamassa synonymiaa. Grinsteadin (2000: 42) aineiston esimerkkien perusteella tämä koetin yhdistää kuitenkin pikemmin määritelmän ja termin kuin synonyymiset termit. Kirjoittaja antaa tällöin tekstissä ensin määritelmän, sitten koettimen ja viimeiseksi termin. Esimerkiksi ”A business consisting of a large number of shops is called a chain.” Vastaava esimerkki on myös Bowkerin ja Pearsonin aineistossa: ”This transfer of genetic material is called **gene flow**. Among biologists, gene flow” (Bowker & Pearson 2002: 153). Tutkimukseni venäläisessä aineistossa järjestys voi olla myös päinvastainen, jolloin määriteltävän käsitteen nimitys on välittömästi ennen koetinehdokasta ja määritelmä välittömästi koetinehdokkaan jälkeen, kuten esimerkissä 190:

(190) Непотопляемостью судна называется его способность оставаться на плаву после затопления части его внутреннего объема. (Bezop: 46)

Hakuehto *называется* aineistosta *Bezop* tuotti 10 osumaa, joista seitsemässä koetinehdokas on validi määritelmän koetin. Koetinehdokas ei siis ole erityisen produktiivinen mutta roskan vähäisyyden vuoksi se on kelvollinen määritelmän koetin.

**Koetinehdokas – *это*.** Konkordanssihaku ajatusviivan ja pronominin *это* yhdistelmällä – *это* (se on) tuotti aineistosta *Bezop* 15 osumaa, joista seitsemässä koetinehdokas toimii määritelmän koettimena. Esimerkissä 191 määriteltävän käsitteen nimitys *танкер* (säiliöalus) on välittömästi ennen koetinehdokasta. Koetinehdokkaan jälkeen on sisältö-määritelmä, jossa on yläkäsitteen nimitys ja erottava piirre:

- (191) Танкер – это судно, предназначенное для перевозки жидких грузов, имеющих свободную поверхность. (Bezop: 98)

Säiliöalus määritellään alukseksi, joka on tarkoitettu sellaisten nestemäisten lastien kuljetukseen, joilla on vapaa nestepinta. Neljässä osumassa ei ole määritelmää mutta niissä selitetään käsitteiden välistä eroa, joka on terminologisesti kiinnostava tieto:

- (192) Взрыв и пожар – это явления одной природы, но протекающие по-разному во времени. (Bezop: 102)

Esimerkissä 192 verrataan käsitteitä *взрыв* (räjähdys) ja *пожар* (palo), joiden erottava piirre on tapahtumaan kuluva aika. Ajatusviivan ja pronominin yhdistelmä, – *это*, osoittautui kelvolliseksi määritelmän koettimeksi, sillä noin joka toisessa osumassa on määritelmä. Lisäksi osassa osumista annetaan käsitteistä muuta terminologisesti kiinnostavaa tietoa.

Aiemmissa tutkimuksissa (Antonova 1984: 111; Šelov 1998: 41, 47; Šilova 2005: 12) saatujen tulosten perusteella tein konkordanssihaut myös hakuehdoilla *подразумевается* (tarkoitetaan), *определ\** (määritelmä; määritellä), *то, что* (se, mikä); *то, для чего* (se, mitä varten); *котор\** (joka; mikä). Konkordanssihakua aineistosta Bezop hakuehdolla *подразумевается* ei tuottanut yhtään osumaa. Sen sijaan hakuehto *определ\** tuotti 99 osumaa. Huomattava osa osumista oli substantiivisia määrittävän partisiippi-muodon *определенный* (määrätty; tietty) eri taivutusmuotoja. Osumien määrän vähentämiseksi hakuehtoa on rajattava. Hakuehto *определени\** tuotti 17 osumaa, joista kuitenkin vain kolme on valideja. Verbin *определять* (määritellä) yksikön 3. persoonan muoto *определяет* tuotti 11 osumaa, joista vain kahdessa verbi toimii määritelmän koettimena:

- (193) Конвенция СОЛАС-74 определяет предельную линию погружения как линию, проведенную по борту не ближе чем в 76 мм под верхней поверхностью палубы переборок. (Bezop: 48)

Esimerkissä 193 annetaan Solas-sopimuksen määritelmä *upporajalle*. Määritelmä on sisältö-määritelmä, sillä siinä on annettu yläkäsitteen nimitys ja erottava piirre. Muissa osumissa puhutaan ennemminkin määrittelystä yleensä tai esimerkiksi tason tai ajan määrittämisestä kuin käsitteiden määritelmästä:

- (194) На сегодняшний день универсального определения понятия аварии не существует. (Bezop: 26)

- (195) Поскольку большинство судовых устройств или систем так или иначе связаны с обеспечением безопасности судна и экипажа, то, следовательно, уровень безопасности определяется надежностью этих систем. (Bezop: 9)

Esimerkissä 194 ei määritellä *haverin* (авария) käsitettä, vaan ainoastaan todetaan, että universaalista määritelmää ei ole. Esimerkissä 195 ei anneta *turvallisuuden* määritelmää, mutta kerrotaan, mistä turvallisuuden taso riippuu.

Aiemmissa tutkimuksissa on venäläisten määritelmän koettimien luetteloon otettu pronomini-ilmaukset *то, что*; *то, для чего* ja pronomini *который*. Konkordanssihakua hakuehdolla *то, что* aineistosta Bezop tuotti 13 osumaa. Yhdessäkään niistä koetinedokas ei toimi määritelmän koettimena. Hakuehto *то, для чего* ei tuottanut yhtään osumaa aineistosta Bezop. Hakuehto *котор\** tuotti 189 osumaa, mikä on liian suuri



määrä osumia, jotta niiden tarkastelu manuaalisesti olisi mielekästä. Nämä koetinehdokkaat ovat muissa tutkimuksissa ja muunlaisella aineistolla osoittautuneet hyviksi koettimiksi, mutta tämän tutkimuksen aineistolla ne eivät sitä ole. Jälleen on todettava, että koettimet ovat kirjoittaja- ja aineistokohtaisia.

**Yhteenveto.** Koetinhaku tuotti aineistosta yhden erityisen vahvan määritelmän venäläisen verbikoettimen. Se on verbi *представлять собой*, jonka tuottamista osumista kaikki yhtä lukuun ottamatta ovat valideja tapauksia ja 16 validista tapauksesta kymmenessä on sisältömääritelmän muotovaatimukset täyttävä määritelmä. Verbi *ymmärtää* ei suomalaisella aineistolla osoittautunut produktiiviseksi määritelmän koettimeksi. Sen venäläiset lähivastineet verbimuodot *понимается* ja *понимают* olivat melko produktiivisia ja tuottivat vain vähän roskaa. Yhdessä verbimuodon *называется* kanssa ne ovat produktiivisimmat muodollisen määritelmän venäläiset koettimet, sillä kaikissa niiden tuottamissa valideissa tapauksissa määritelmässä on ainakin jonkin tason yläkäsite ja erottava piirre. Venäläinen koetin, jolle ei ole vastinetta englannissa tai suomessa, on ajatusviivan ja pronominin *это* yhdistelmä. Se tuotti seitsemän määritelmää, jotka kaikki ovat sisältömääritelmiä. Sen sijaan sulkeet ei ole hyvä määritelmän koetin, sillä ne tuottavat runsaasti roskaa. Venäläisellä aineistolla sulkeissa on yleensä termi ja määritelmään kuuluva käsitepiirre ilman yläkäsitettä on ennen sulkeita.

#### 7.5.4 Geneerisen käsitesuhteen ylä- ja alakäsitteen venäläiset koettimet

Koettimien tutkimuksessa on tutkijoiden mielenkiinto kohdistunut useimmiten geneerisen suhteen koettimiin. Tämä on ymmärrettävää: pidetäänhän geneeristä suhdetta käsitteiden välisistä suhteista tärkeimpänä. Se on yleensä lähtökohta sanastotyössä, kun käsitteiden väliset suhteet halutaan selvittää. Geneerisen suhteen koettimien luettelo niissä kielissä, joissa koettimia ylipäätään on tutkittu, on varsin pitkä, sillä luettelon pituus korreloi suoraan tutkijoiden mielenkiinnon kanssa. Geneerisen suhteen venäläisiä koettimia ovat tutkineet ainakin Antonova (1984: 111), Piotrovskij ym. (1981: 35), Šelov (1998: 41) ja Šilova (2005: 12). He esittävät luettelon, jossa on pääasiassa geneerisen suhteen koettimina toimivia verbejä, mutta myös yksi välimerkki eli ajatusviiva ja yksi pronomini, *это* (tämä). Geneerisen suhteen koettimina toimivista verbeistä *olla*-verbiä pidetään yleisesti hyvänä geneerisen suhteen koettimina kielestä riippumatta. Juuri *olla*-verbin yksikön 3. persoonan indikatiivin preesensmuodot *is* ja *are* toimivat koettimina englannissa. Suomen kielessä verbi on kuitenkin ongelmallinen, kuten totesin aiemmin jaksossa 7.4.4. Venäjässä ei preesens-muotoista kopulaa käytetä juuri lainkaan ja *быть*-verbin suplettiivinen preesensmuoto *есть* (on) esiintyy lähes pelkästään eksistentiaalilauseissa. Konkordanssihaku hakuehdolla *есть* ei tuottanutkaan artikkelista *Eskos* yhtään osumaa. *Bezop*-aineistosta hakuehto tuotti 29 osumaa, mutta yhdessäkään niistä se ei ilmaise geneeristä suhdetta. Verbimuodon *есть* sijasta predikatiivin kanssa voidaan käyttää esimerkiksi verbin *являться* (*явиться*) (*olla*) yksikön tai monikon 3. persoonan indikatiivin preesensmuotoja *является* tai *являются*. Konkordanssihaun tein hakuehdolla *является/являются*.

Koetinehdokkaiden tunnistamiseksi tehty konkordanssihaku tunnetuilla termeillä artikkelista *Eskos* tuotti mahdollisiksi geneerisen suhteen venäläisiksi koettimiksi sulkeet, adjektiivit *другой* (muu) ja *прочий* (muu), lyhenteen *и т.д.* (jne.) sekä sanan *например* (esimerkiksi). Aineistosta *Bezop* tekemäni manuaalinen haku tuotti koetinehdokkaiden listaan lyhenteen *и др.*, verbit *относиться* ja *отнести*, ilmauksen *в качестве* sekä substantiivit *вид*, *тип* ja *форма*. Koetinehdokkaista muodostettujen hakuehtojen tuottamat tulokset ovat taulukossa 51.

TAULUKKO 51. Geneerisen käsitesuhteen terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
являться (явиться)	<i>является/являются</i>	157	28 (18)	18	heikko
sulkeet	<i>(*</i>	370	46	12	heikko
другой	<i>други*</i>	63	26	41	hyvä
и т.д.	<i>и т.д.</i>	32	7	22	heikko
и др.	<i>и др.</i>	16	6	38	hyvä
например	<i>например</i>	50	12	24	heikko
относиться	<i>относятся</i>	10	6	60	hyvä
отнести	<i>отнести</i>	4	3	75	hyvä
в качестве	<i>в качестве</i>	25	10	40	hyvä
вид	<i>вид*</i>	64	29	45	hyvä
тип	<i>тип*</i>	36	12	33	heikko
форма	<i>форм*</i>	57	4	7	heikko

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *являться*.** Tunnettujen termien ympäristön analyysi *Eskos*-artikkelista tuotti yhden tapauksen, jossa verbimuoto *является* (on) toimii geneerisen suhteen koettimena:

- (196) Примером региональной ЕСКУС является проект единой системы залива Петра Великого (ЕСПВ), - - (Eskos: 18)

Esimerkissä 196 annetaan esimerkki käsitteen *региональная ЕСКУС* (alueellinen VTMISS) alaan kuuluvasta yksilökäsitteestä *единая система залива Петра Великого (ЕСПВ)*, joka on yksilökäsitteenä käsitejärjestelmän alimmalla tasolla. Aineistosta *Bezop* konkordanssihakua hakuehdolla *является/являются* tuotti 139 osumaa, joista 27 osumassa verbimuoto *является* (on) tai *являются* (ovat) toimii geneerisen suhteen koettimena. Viidessä validissa osumassa verbi on monikkomuodossa, kuten esimerkissä 197, ja 22 osumassa yksikkömuodossa, kuten esimerkissä 198:

- (197) Основными формами подготовки экипажа являются занятия, тренировки, учения и самостоятельная подготовка, причем последняя является основной формой подготовки. (Bezop: 66)

- (198) Система углекислотного тушения является наиболее распространенной системой тушения пожаров в грузовых трюмах, МО, кладовых, топливных и масляных танках. (Bezop: 60)

Esimerkissä 197 geneerisen suhteen yläkäsite *форма подготовки экипажа* (miehistön koulutusmuoto) on välittömästi ennen koetinehdokasta *являются* ja alakäsitteet *занятие* (oppitunti), *тренировка* (harjoittelu, harjoitus), *учения* (harjoitukset) ja *самостоятельная подготовка* (itseopiskelu) välittömästi sen jälkeen. Esimerkissä 198 järjestys on päinvastainen, sillä alakäsite *система углекислотного тушения* (CO<sub>2</sub>-

sammutusjärjestelmä) on välittömästi ennen koetinehdokasta ja yläkäsitem *система тушения пожаров* (palonsammutusjärjestelmä) koetinehdokkaan jälkeen. 27 validista tapauksesta 23 tapauksessa geneerisen suhteen ylä- tai alakäsitem on välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen.

Vaikka koetinehdokas *является/являются* toimii artikkelissa *Eskos* vain kerran geneerisen suhteen koettimena, laajemmalla *Bezop*-aineistolla tehdyn haun perusteella se kannattaa hyväksyä geneerisen suhteen koetimeksi, koska valideissa osumissa välittömästi sitä ennen tai sen jälkeen on lähes aina termi. Se ei kuitenkaan ole vahva koetin, koska haku ilman tunnettua termiä tuottaa erittäin paljon roskaa.

**Koetinehdokas *sulkeet*.** Koetinehdokkaista sulkeet tuottivat eniten valideja geneerisen suhteen osumia. Venäläisestä aineistosta tehty konkordanssihakuehdolla (\* tuotti 46 sellaista tulosta, jossa ennen sulkeita on yläkäsitteen nimitys ja sulkeissa vähintään yksi geneerisen alakäsitteen nimitys. Aineistosta riippumatta sulkeissa luetellut käsitteet ovat lähes aina geneerisen käsitejärjestelmän lähikäsitteitä eivätkä koostumussuhteisen käsitejärjestelmän lähikäsitteitä. 39 tuloksessa yläkäsitem on välittömästi ennen sulkeita:

- (199) Балкер – специализированное судно, предназначенное для перевозки массовых грузов (руды, удобрений, зерна и т.п.) навалом. (Bezop: 109)

Esimerkissä 199 ennen sulkeita on termi *массовый груз* (kuiva irtolasti) ja sulkeissa ovat alakäsitteiden nimitykset *руда* (malmi), *удобрение* (lannoite) ja *зерно* (vilja). Kolmessa tuloksessa yläkäsitem on ennen sulkeita, mutta sen ja sulkeiden välissä on käsitteeseen kuulumattomia sanoja. Esimerkissä 200 yläkäsitteen ja sen geneeristen alakäsitteiden välissä on yläkäsitteen *огнетушащее вещество* (sammute) määritelmä:

- (200) Огнетушащие вещества — огнегасительные средства пожарных систем (пенообразователи, порошки, хладоны, газы). (Bezop: 62)

Hakuehto tuotti myös kaksi osumaa, joissa sulkeissa oleva käsite tai käsitteet ovat koostumussuhteessa käsitteeseen, joka on ilmaistu ennen sulkeita:

- (201) Эти системы определяют только наличие или отсутствие воды в льялах и колодцах путем световой и звуковой индикации на пульте в центральном посту управления (в МО). (Bezop: 50)

Esimerkissä 201 sulkeissa on yläkäsitettä *машинное отделение* (konehuone) nimeävä lyhennetermi *МО*. Sen koostumussuhteinen alakäsitem on konehuoneessa sijaitseva *центральный пост управления* (konevalvomo), jota nimeävä termi on tekstissä välittömästi ennen sulkeita.

**Koetinehdokkaat *другой* ja *прочий*.** Geneerisen suhteen venäläisten koettimien poimintaa varten tehty konkordanssihakuehdolla tunnetuilla termeillä artikkelista *Eskos* tuotti sulkeiden jälkeen toiseksi eniten osumia adjektiivista *другой* (muu) muodostetulla hakuehdolla *други*\*. Yhdeksästä osumasta viisi on valideja geneerisen suhteen kontekstiesimerkkejä. Adjektiivin *прочий* monikkomuoto *прочие* (muut) tuotti hakuehdolla *прочи*\* kaksi tulosta, joista yhdessä adjektiivi toimii geneerisen suhteen koettimena. Esi-merkeistä päätellen adjektiivit *другие* ja *прочие* toimivat samalla tavoin:

- (202) Для доступа к информационным ресурсам региональной и бассейновой информационных систем государственных органов, осуществляющих надзор за судоходст-

вом (МО, ФПС, ГТК), и других участников транспортного процесса (судовладельцев, агентов, портовиков и т.д.) целесообразно использовать глобальные компьютерные сети, - - (Eskos: 19)

- (203) Информационная локальная ЕСКУС интегрирует базы данных службы капитана порта (СКП), СУДС, ССС или АЙС [sic], обеспечивает обмен информацией между судовыми и наземными базами данных, обеспечивает доступ к информации для местных государственных органов, осуществляющих надзор за судоходством (МО, ФПС, ГТК), и для прочих участников транспортного процесса (судовладельцев, агентов, портовиков и т.д.). (Eskos: 18)

Esimerkissä 202 geneerisen suhteen koettimena toimiva adjektiivi *другие* on tekstissä välittömästi ennen yläkäsitettä nimeävää termiä *участник транспортного процесса* (kuljetusketjun lenkki). Koetinehdokas kertoo lukijalle, että vähintään yksi alakäsite on mainittu tekstissä aiemmin. Tämä alakäsite on *государственный орган, осуществляющий надзор за судоходством* (alusliikennettä valvova viranomainen). Osa muista alakäsitteistä on annettu sulkeissa yläkäsitteen jälkeen. Lyhenne *и т.д.* (jne.) alakäsitteiden luettelon perässä osoittaa, että luettelo ei ole täydellinen eli kaikkia alakäsitteitä ei ole lueteltu.

Esimerkissä 203 alakäsitteen nimityksenä on *местный государственный орган, осуществляющий надзор за судоходством*. Nimityksellä viitataan ilmeisesti samaan käsitteeseen kuin edellisen esimerkin nimityksellä, koska kummankin jälkeen sulkeissa on sama luettelo merenkulkua valvovista viranomaisista. Kaikissa adjektiivien *другие* ja *прочие* valideissa osumissa yläkäsitteen nimitys on välittömästi koetinehdokkaan jälkeen ja ainakin yhden alakäsitteen nimitys on välittömästi ennen koetinehdokasta. Kahdessa osumassa on mainittu vain yksi alakäsite, mutta koetinehdokas ilmaisee lukijalle, että muitakin alakäsitteitä on olemassa:

- (204) - - предоставления информации государственным структурам, обеспечивающим контроль над судоходством, и другим участникам транспортного процесса. (Eskos: 16)

Esimerkissä 204 yläkäsite on jälleen *участник транспортного процесса*. Esimerkissä nimetään alakäsite *государственная структура, обеспечивающая контроль над судоходством* (alusliikennettä valvova viranomainen), joka saattaa viitata samaan käsitteeseen kuin edellisten esimerkkien *государственный орган, осуществляющий надзор за судоходством* ja *местный государственный орган, осуществляющий надзор за судоходством*, mutta tässä esimerkissä ei ole kerrottu, mitä viranomaisia laivaliikennettä valvovilla viranomaisilla tarkoitetaan.

- (205) - - в том числе торгового судоходства, промышленного рыболовства, добыча полезных ископаемых и другие виды хозяйственного использования морей и побережья. (Eskos: 15)

Esimerkissä 205 alakäsitteet *торговое судоходство* (kauppamerenkulku), *промышленное рыболовство* (ammattikalastus) ja *добыча полезных ископаемых* (kaivostointiminta) on lueteltu ennen koetinehdokasta *и другие* ja yläkäsitteen nimitys *хозяйственное использование морей и побережья* (merten ja rannikoiden talouskäyttö) on välittömästi koetinehdokkaan jälkeen. Koetinehdokasta vahvistaa sana *вид*, joka osoittaa, että alakäsitteet ovat yläkäsitteen ”lajeja” eli geneerisessä suhteessa yläkäsitteeseen.

*Bezop*-aineistosta hakuehdolla *δpyzu*\* tehty konkordanssihaku tuotti 54 osumaa, joista 21 on valideja tapauksia koetinehdokkaan *δpyzue* toiminnasta geneerisen suhteen koettimena:

- (206) Применение пара для тушения горящего угля, жмыхов, хлопка и других волокнистых грузов малоэффективно. (*Bezop*: 80)

Esimerkissä 206 alakäsitteiden nimitykset *уголь* (hiili), *жмых* (öljyakku) ja *хлопок* (puuvilla) ovat ennen koetinta *δpyzux* ja yläkäsitteen nimitys *волокнистый груз* (kuitumainen lasti) välittömästi koettimen jälkeen. Yläkäsitteen nimitys on välittömästi koettimen jälkeen kaikissa muissa valideissa osumissa paitsi viidessä elliptisessä osumassa, joista yläkäsitteen nimitys puuttuu:

- (207) Применяются датчики биметаллические, плавкие, термодифференциальные, ионизационные и другие, однако наибольшее распространение получили датчики с плавкой вставкой и биметаллические. (*Bezop*: 57)

Esimerkissä 207 yläkäsitteen nimitys *датчик* (ilmaisín) puuttuu koettimen yhteydestä. Myös alakäsitteiden nimitykset ovat elliptisiä ensimmäistä lukuun ottamatta.

Hakuehto *δpyzu*\* tuotti *Bezop*-aineistosta viisi osumaa, joissa on sanaliitto *δpyzumi* *словами* (toisin sanoen), ja yksittäisiä osumia, joissa pronominiin liittyy yleiskielinen substantiivi, kuten esimerkiksi *средство* (väline, keino), *способ* (keino), *возможность* (mahdollisuus), *аспект* (piirre) tai *причина* (syy). Näitä osumia ei ole laskettu validien osumien joukkoon.

Konkordanssihaulla artikkelista *Eskos* poimittujen tapausten perusteella geneerisen suhteen adjektiivikoettimet *δpyzue* ja *прочие* toimivat hyvin samalla tavalla, mutta adjektiivi *δpyzue* on produktiivisempi. Adjektiivi *прочие* ei ollut produktiivinen myöskään aineistossa *Bezop*, sillä hakuehto *прочу*\* tuotti vain viisi osumaa, joista yksikään ei ole validi tapaus koetinehdokkaan toiminnasta geneerisen suhteen koettimena. Osumista neljässä on sanaliitto *наряду с прочим* (muun muassa), joka on pikemmin täydentävä lausahdus kuin terminologinen koetin.

**Koetinehdokas u m.δ.** Konkordanssihaku hakuehdolla *u m.δ.* (jne.) artikkelista *Eskos* tuotti neljä osumaa, joista kolme on validia esimerkkiä lyhenteen toiminnasta geneerisen suhteen koettimena. Kaikissa näissä osumissa lyhenne on geneeristen alakäsitteiden luettelon lopussa osoittamassa, että luettelo ei ole täydellinen. Voidaan tietysti kysyä, ilmaiseeko lyhenne nimenomaan geneeristä suhdetta, mutta koska sen avulla pystytään paikallistamaan alakäsitteiden luettelo, on yläkäsitteen nimityskin todennäköisesti lähi-kontekstissa. Toisaalta kaikissa valideissa osumissa on kaksi muutakin geneerisen suhteen koetinta, sulkeet ja adjektiivi *δpyzue* tai *прочие*, joten näistä koettimista voidaan valita yksi, joka tuottaa eniten valideja osumia suhteessa kaikkien osumien määrään. Näin mitattuna parhaan hyödyn tuottaa adjektiivi *δpyzue*, joka tuotti artikkelista *Eskos* hakuehdolla *u δpyzu*\* seitsemän osumaa, joista viisi on valideja.

Aineistosta *Bezop* hakuehto *u m.δ.* antoi 28 osumaa, joista neljässä lyhenne toimii geneerisen suhteen koettimena. Jokaisessa osumassa lyhennettä edeltää luettelo, mutta niissä luetellaan pikemmin syitä tai toimenpiteitä kuin geneerisiä alakäsitteitä. Lisäksi mahdollinen yläkäsitteen nimitys on tekstissä niin kaukana ennen lyhennettä, että se ei tulostu konkordanssilistaan oletusasetuksella, jonka mukaan listaan tulostuu kuusi sanaa ennen ja kuusi sanaa jälkeen hakulausekkeen. Aineistosta *Bezop* tehdyn konkordanssihaun perusteella lyhenne *u m.δ* osoittaa ainoastaan, että edeltävä luettelo ei ole täydelli-



nen. Lyhenteen *u m.δ* erilainen käyttö aineistoissa saattaa merkitä sitä, että koetinehdokas on kirjoittajakohtainen.

**Koetinehdokas *u δp*.** Manuaalinen koetinhaku aineistosta *Bezop* tuotti lyhenteen *u m.δ*. lisäksi toisenkin mahdollisen geneerisen suhteen lyhennekoettimen, *u δp*. (ja muut) Hakuehto *u δp*. aineistosta *Bezop* tuotti 16 osumaa, joissa kaikissa lyhennettä edeltää käsitteiden nimitysten luettelo. Kuudessa tapauksessa luetteloa edeltää geneerisen yläkäsitteen nimitys, joka on neljässä tapauksessa välittömästi ennen alakäsitteiden nimityksiä, kuten esimerkissä 208 yläkäsitteen nimitys *инертный газ* (inerttikaasu):

- (208) Инертные газы (азот, аргон, гелий, дымовые или отработанные газы и др.) применяются в основном с целью понижения концентрации кислорода в очаге горения. (Bezop: 81)

Kahdessa tapauksessa luetteloa edeltää koostumussuhteisen yläkäsitteen nimitys, ja neljässä tapauksessa yläkäsitettä ei ole ilmaistu tai se ei ole merenkulun käsite.

**Koetinehdokas *например*.** Sulkujen, adjektiivien *другие* ja *прочие* sekä lyhenteiden *u m.δ*. ja *u δp*. lisäksi mahdollinen geneerisen suhteen koetin on tunnetuilla termeillä tehdyn koetinhaun perusteella sana *например* (esimerkiksi). Konkordanssihakua hakuehdolla *например* aineistosta *Bezop* tuotti 49 osumaa, joista 12 osumassa sana toimii geneerisen suhteen koettimena. Alakäsitteen nimitys on kaikissa valideissa osumissa välittömästi koetinehdokkaan jälkeen. Joka toisessa osumassa koetinehdokas on sulkeissa sitä seuraavan alakäsitteen nimityksen tai alakäsitteiden nimitysten kanssa:

- (209) Эта система позволяет в особо опасных помещениях (например, помещение электромоторов) поддерживать избыточное давление воздуха, - - (Bezop: 105)

Esimerkissä 209 yläkäsitteen nimitys *особо опасное помещение* (erityisen vaarallinen tila) on välittömästi ennen sulkeissa olevaa koetinehdokasta *например* ja alakäsitteen nimitys *помещение электромоторов* (akkuhuone) välittömästi koetinehdokkaan jälkeen.

- (210) Остальные 75% аварий принято относить к субъективным факторам, таким, например, как; пренебрежение мерами безопасности; отсутствие чувства ответственности у членов экипажа; недостаточная профессиональная подготовка; нарушение требований нормативных документов; слабая организация вахтенной службы; чрезмерная усталость экипажа; нечеткая организация работы аварийно-спасательных служб и т.д. (Bezop: 25)

Esimerkissä 210 koetinehdokas ei ole sulkeissa, mutta sen sekä yläkäsitteen nimityksen *субъективный фактор* (inhimillinen tekijä) ja alakäsitteiden nimitysten luettelon välissä on sana *таким* (sellaisiin) tai sana *как* (kuten). Kahdessa osumassa yläkäsitteen nimitys on konkordanssilistan tulostusalueen ulkopuolella, jos tulostuksessa käytetään ohjelman perusasetuksia. Sana *например* on tämän tutkimuksen aineistolla kyllin produktiivinen, jotta se voidaan hyväksyä geneerisen käsitesuhteen venäläiseksi koettimeksi, mutta paljon roskaa tuottavana se toimii koettimena parhaiten yhdessä tunnetun termin kanssa.

Olen edellä käsitellyt tunnetuilla termeillä artikkelista *Eskos* tehdyn konkordanssinhaun tuottamia geneerisen käsitesuhteen venäläisiä koetinehdokkaita, joita ovat sulkeet, adjektiivit *другой* (muu) ja *прочий* (muu), lyhenteet *u m.δ*. (jne.) ja *u δp*. (ym.) sekä sana *например* (esimerkiksi). Aineistosta *Bezop* tehty manuaalinen koetinhaku tuotti



muiksi mahdollisiksi venäläisiksi geneerisen suhteen koettimiksi verbit *относиться* (*относятся*) (kuulua johonkin), *отнести* (lukea kuuluvaksi johonkin), *бывать* (*бывает*) (olla), substantiivit *вид* (laji), *тип* (tyyppi), *форма* (muoto), partikkelin *как* (kuten) ja ilmauksen *в качестве* (jonakin). Partikkeli *как* tuotti konkordanssihaussa 224 osumaa. Määrä on niin suuri, että konkordanssilistan manuaalinen käsittely ei ole mielekästä. Tästä syystä en tutki tarkemmin partikkelin *как* toimintaa koettimena. Sen sijaan selostan seuraavaksi muilla manuaalisen koetinhaun tuottamilla koetinehdokkailla *Bezop*-aineistosta tehtyjen konkordanssihakujen tuloksia.

**Koetinehdokas *относиться*.** Aineistosta *Bezop* tehty konkordanssihaku verbin *относиться* monikon 3. persoonan muodolla *относятся* (kuuluvat johonkin) tuotti 10 osumaa, joista kuudessa koetinehdokas osoittaa geneeristä suhdetta. Viidessä validissa osumassa joko yläkäsitteen tai alakäsitteen nimitys on välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen:

- (211) К таким насосам относятся поршневые, мембранные, центробежные с системой удаления воздуха. (*Bezop*: 50)

Esimerkissä 211 yläkäsitteen nimitys *насос* (pumppu) on välittömästi ennen koetinehdokasta ja alakäsitteiden nimitykset ovat välittömästi koetinehdokkaan jälkeen. Kahdessa tapauksessa järjestys on päinvastainen:

- (212) К средствам противопожарной защиты, помимо конструктивных мер и сигнализации, относятся следующие системы пожаротушения: 1) водопожарная; 2) спринклерная; 3) водораспыления; 4) водяных завес; 5) водяного орошения; 6) паротушения; 7) углекислотного тушения; 8) инертных газов; 9) тушения хладагентами; 10) пенотушения; 11) порошкового тушения. Все системы пожаротушения можно условно разделить по принципу действия гасящего агента. Первые шесть систем относятся к водогасящим, 7–10-я – имеют химически гасящий агент, 10–11-я – механически гасящие. (*Bezop*: 57–58)

Esimerkissä 212 alakäsitteiden nimitykset ovat paljon ennen koetinehdokasta. Sen sijaan yläkäsitteen nimitys *водогасящая система пожаротушения* (vesisammutusjärjestelmä) on välittömästi koetinehdokkaan ja siihen liittyvän preposition *к* jälkeen. Kaikissa osumissa verbiin *относиться* liittyy prepositio *к*, mutta sen paikka tekstissä vaihtelee eikä sen avulla voi erottaa kelvollisia osumia muista, koska prepositio näyttää liittyvän verbimuotoon *относятся* myös muissa kuin valideissa osumissa.

**Koetinehdokas *отнести*.** Manuaalinen koetinhaku osoitti, että myös verbi *отнести* (lukea johonkin kuuluvaksi) on liitettävä venäläisten geneerisen suhteen koettimien luetteloon, sillä hakuheito *отнести* tuotti aineistossa *Bezop* neljä osumaa, joista kolmessa verbi toimii geneerisen suhteen koettimena. Perusmuodossa olevaan verbiin liittyy kaikissa valideissa tapauksissa jokin adverbi, kuten esimerkiksi predikaatiivinen adverbi *можно* (voi) tai *нужно* (pitää, täytyy) tai adverbi *правоммерно* (oikeutetusti):

- (213) Технологические аварии связаны с грузовой функцией судна. К ним можно отнести: нарушение прочности корпуса судна, его устройств, потерю остойчивости судна, подвижку груза, возникновение пожара, попадание воды в корпус судна, образование взрывоопасных застойных зон паров и газов в судовых грузовых помещениях и т.д. (*Bezop*: 37–38)

Kolmas geneerisen suhteen mahdollinen verbikoetin *бывать*-verbin yksikön 3. persoonan muoto *бывает* (on usein, voi olla) tuotti yhdeksän osumaa, joista kolmessa koetinhdokas toimii geneerisen suhteen koettimena:

(214) Углекислотная система бывает низкого и высокого давления. (Bezop: 60)

Esimerkissä 214 yläkäsitteen nimitys *углекислотная система* (CO<sub>2</sub>-sammutusjärjestelmä) on välittömästi koetinhdokkaan *бывает* edellä ja elliptiset alakäsitteiden nimitykset (*углекислотная система*) *низкого давления* (paineistamaton CO<sub>2</sub>-sammutusjärjestelmä) ja (*углекислотная система*) *высокого давления* (paineistettu CO<sub>2</sub>-sammutusjärjestelmä) ovat välittömästi koetinhdokkaan jälkeen, kuten muisakin valideissa osumissa. Verbimuoto *бывает* toimiikin koettimena vain yhdessä tunnetun termin kanssa.

**Koetinhdokas в качестве.** Geneerisen suhteen mahdollisilla substantiivikoettimilla *вид* (laji), *тип* (tyyppi) ja *форма* (muoto) sekä ilmauksella *в качестве* (jonakin) tein konkordanssihaun hakuehdoilla *в качестве*, *вид\**, *тип\** ja *форм\**. Hakuehto *в качестве* tuotti Bezop-aineistosta 25 osumaa. Niistä 10 on sellaisia, joissa koetinhdokkaan voi ajatella osoittavan geneerisen käsitesuhteen yläkäsitteen ja alakäsitteen välistä suhdetta:

(215) Так, вода в качестве гасящего агента понижает температуру; - - (Bezop: 80)

Esimerkissä 215 yläkäsitteen nimitys *гасящий агент* (sammute) on välittömästi koetinhdokkaan jälkeen ja alakäsitteen nimitys *вода* (vesi) välittömästi ennen koetinhdokasta. Koetinhdokkaan etu on, että yläkäsitteen nimitys on kaikissa valideissa osumissa välittömästi koetinhdokkaan jälkeen.

**Koetinhdokas вид.** Geneerisen suhteen mahdollinen substantiivikoetin *вид* (laji) tuotti hakuehdolla *вид\** 64 osumaa, joista 29 on validia. 20 validissa osumassa koetinhdokasta seuraa välittömästi yläkäsitteen nimitys:

(216) Но самый опасный вид взрыва, конечно, химический – мгновенное горение, для чего должны быть соответствующая концентрация газа в воздухе и источник воспламенения. (Bezop: 104)

Esimerkissä 216 yläkäsitteen nimitys *взрыв* (räjähdys) on välittömästi koetinhdokkaan jälkeen ja alakäsitteen elliptinen nimitys *химический (взрыв)* (kemiallinen räjähdys) on yhden sanan päässä yläkäsitteen nimityksen jälkeen. Lauseen syntaktinen rakenne on tällainen, koska yläkäsitettä nimeävä termi on sanan *вид* genetiiviattribuutti ja näin ollen sen paikka venäjän kielessä on pääsanana jäljessä. Niissä osumissa, joissa koetinhdokas yhdistyy yläkäsitteen nimitykseen, alakäsitteen nimitys on yleensä kaukana tästä. Vain viidessä tapauksessa alakäsitteen nimitys on niin lähellä yläkäsitteen nimitystä, että se on tulostunut konkordanssilistaan. Kolmessa validissa osumassa alakäsitteen nimitys on välittömästi koetinhdokkaan jälkeen. Esimerkissä 217 välittömästi koetinhdokkaan jälkeen ovat alakäsitteiden nimitykset *переборка* (laipio) ja *палуба* (kansi):

(217) Эти перекрытия могут быть вертикальными (в виде переборок) и горизонтальными (в виде палуб). (Bezop: 54)

Hakua on mahdollista tarkentaa sulkemalla osumalistasta pois kaikki osumat, joissa koetinehdokas on muodossa *в виду*. Seitsemässä osumassa on ilmaus *следует иметь в виду* (on pidettävä mielessä). Substantiivi *вид* ei ole kovin hyvä koetin, koska noin puolet osumista on roskaa. Valideissa osumissa todennäköisesti vain geneerisen suhteen yläkäsitteen nimitys on tulostunut konkordanssilistaan mutta alakäsitteiden nimitykset ovat tulostusalueen ulkopuolella.

**Koetinehdokas *mun*.** Konkordanssihaku aineistosta *Bezop* substantiivista *mun* (tyyppi) muodostetulla hakuehdolla *mun\** tuotti 36 osumaa, joista 12:ssa koetinehdokkaan kontekstissa on sekä ylä- että alakäsitteen nimitys. Neljässä tapauksessa yläkäsitteen nimitys on välittömästi koetinehdokkaan jälkeen, kuten termi *пластырь* (vuotomatto) esi-merkissä 218:

- (218) - - на судах применяются три основных типа пластырей, указанных Регистром. Кольчужный пластырь применяют на судах неограниченного района плавания длиной более 150 м, кроме танкеров (рис. 3.1). (Bezop: 51)

Neljässä osumassa koetinehdokasta seuraa erisnimi, esimerkiksi aluksen nimi, jota käytetään nimikkeenä kokonaiselle sarjalle samanlaisia aluksia:

- (219) По этой же причине на учебных парусных судах типа «Дружба» шлюпки закрывают обзор с мостика. (Bezop: 12)

Koetinehdokas *mun* toimii hyvin samalla tavoin kuin koetinehdokas *вид*. Myös sen tuottama terminologinen tieto on haettava kontekstista, sillä samoin kuin substantiivin *вид* myös substantiivin *mun* tuottamissa osumissa alakäsitteen nimitys tai alakäsitteiden nimitykset ovat tekstissä kaukana koetinehdokkaasta. Esimerkissä 218 ensimmäinen alakäsitteen nimitys on eri virkkeessä kuin koetinehdokas ja yläkäsitteen nimitys. Tosin tässä esimerkissä koetinehdokkaan ja ensimmäisen alakäsitteen nimityksen välissä on yläkäsitteen nimityksen lisäksi vain kaksi sanaa.

**Koetinehdokas *форма*.** Substantiivi *форма* (muoto) käyttäytyy koettimena hyvin samalla tavoin kuin substantiivit *вид* ja *mun*, sillä myös sen tuottamissa valideissa osumissa yläkäsitteen nimitys on välittömästi koetinehdokkaan jälkeen. Sen tuottamien terminologisesti kiinnostavaa tietoa sisältävien osumien määrä aineistossani on kuitenkin pieni. Tein ensin konkordanssihaun hakuehdolla *форма\**, joka tuotti 57 osumaa, mutta niistä vain neljässä on geneerinen käsitesuhde:

- (220) Борьба с паром является одной из форм борьбы за живучесть, - - (Bezop: 78)

Esimerkissä 220 yläkäsitteen nimitys *борьба за живучесть* (pelastaminen) on välittömästi koetinehdokkaan jälkeen, mutta 53 osumassa ei ole sen paremmin yläkäsitteen kuin alakäsitteenkään nimitystä. Validien osumien mukaan muotoiltu tarkennettu hakuehto *форма/форм/формой/формами* tuottaa 11 osumaa, joten roskaosumien osuus vähenee huomattavasti. Aineistoni on kuitenkin liian pieni, jotta sen perusteella voisi päätellä, mitkä taivutusmuodot hakuehtoon kannattaa valita. On todennäköistä, että substantiivi *форма* voi koettimena toimiessaan olla tarkennetun hakuehdon taivutusmuotojen lisäksi muissakin muodoissa.

**Yhteenveto.** Tutkimukseni ei yllättäen tuottanut yhtään vahvaa geneerisen käsitesuhteen ylä- ja alakäsitteen venäläistä koetinta. Koetinehdokkaista sulkeet tuottivat eniten

osumia. Niiden joukossa on kyllä runsaasti valideja osumia mutta myös paljon roskaa. Parhaiksi koettimiksi osoittautuivat adjektiivin *другой* monikko *другие* ja substantiivi *вид*, jotka ovat produktiivisia ja lisäksi lähes puolet osumista on valideja. Substantiivi *вид* samoin kuin substantiivit *мун* ja *форма* sekä ilmaus *в качестве* toimivat koettimina parhaiten yhdessä termin kanssa, sillä tekstissä yläkäsitteen nimitys seuraa tavallisesti välittömästi koettimen jälkeen. Tarkimpia ovat verbikoettimet *относятся* ja *отнести*, jotka tuottavat runsaasti valideja osumia suhteessa kaikkien osumien määrään. Ne eivät kuitenkaan ole erityisen produktiivisia. Myös verbi *являться* voidaan hyväksyä koettimien luetteloon, mutta vahva koetin sekään ei ole, koska se tuottaa paljon roskaa. Sana *например* on kelvallinen koetin vain yhdessä termin kanssa, sillä se tuottaa runsaasti osumia, joista kuitenkin vain noin neljännes on valideja.

### 7.5.5 Koostumussuhteen ylä- ja alakäsitteen venäläiset koettimet

Venäläisiä geneerisen suhteen koettimia selvittäneet tutkijat ovat sisällyttäneet tutkimuksiinsa myös koostumussuhteen koettimet (Antonova 1984: 111; Piotrovskij ym. 1981: 35; Šelov 1998: 145; Šilova 2005: 12). Näiden tutkimusten perusteella venäläisiä koostumussuhteen koettimia ovat verbit *состоять* (koostua), *принадлежать* (kuulua) ja *включать* (sisältää), jotka toimivat koettimina yhdessä tiettyjen prepositioiden kanssa, sekä substantiiveista erityisesti substantiivi *часть* (osa), jota Šelov (1998: 145) nimittää vahvaksi koettimeksi. Muita koostumussuhteen substantiivikoettimia voivat olla kokonaisuutta ilmaisevat luokkasanat, kuten esimerkiksi *система* (järjestelmä) ja *комплекс* (kokonaisuus).

Halusin selvittää, saavatko aiempien tutkimusten tulokset tukea tämän tutkimuksen aineistosta. Tarkastelin ensin artikkelista *Eskos* poimittujen termien kontekstia ja kiinnitin huomiota erityisesti niihin konteksteihin, joissa termien nimeämien käsitteiden välillä on koostumussuhde. Mahdollisten koostumussuhteen koettimien käsinpoiminta *Eskos*-artikkelista tuotti viisi koetinehdokasta. Nämä ovat substantiivit *элемент* (osa), *компонент* (osa), *состав* (kokoonpano, koostumus), *структура* (rakenne) ja *часть* (osa). Koetinehdokkaista vain substantiivi *часть* on sama kuin aiemmissa tutkimuksissa.

Manuaalinen koetinehdokkaiden haku aineistosta *Bezop* tuotti mahdollisiksi koostumussuhteen koettimiksi verbien ja prepositioiden yhdistelmät *включает в себя* (sisältää) ja *состоит из* (koostuu). Kolmas aiemmissa tutkimuksissa koostumussuhteen koettimeksi osoittautunut verbi *принадлежать* ei ollut produktiivinen tutkimukseni aineistossa, sillä konkordanssihakua hakuehdolla *принадлежит к* aineistosta *Bezop* ei tuottanut yhtään osumaa. Sen sijaan konkordanssihakua aineistosta *Bezop* hakuehdoilla *состо\** *из* tuotti 17 osumaa, joista 14 osumassa koetinehdokas toimii koostumussuhteen koettimena, ja hakuehto *включает в себя* tuotti kuusi osumaa, joissa kaikissa koetinehdokas on koostumussuhteen koetin. Näillä ja muilla koostumussuhteen koetinehdokkailla tehtyjen konkordanssihakujen tulokset olen koonnut taulukkoon 52.

ТАУЛУККО 52. Koostumussuhteen terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
состоять из	<i>состо* из</i>	17	14	82	vahva
включать	<i>включает в себя</i>	6	6	100	hyvä
элемент	<i>элемент*</i>	57	10	18	heikko
компонент	<i>компонент*</i>	4	3	75	hyvä
часть	<i>часть*</i>	21	5	24	heikko
состав	<i>в состав*</i>	10	5	50	hyvä
структура	<i>структур*</i>	14	3	21	heikko
этап	<i>этап*</i>	12	4	33	hyvä

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *состоять из*.** Hakuehdon *состо\* из* tuottamista 14 validista osumasta 12 osumassa koetinehdokas on muodossa *состоит из* (koostuu) ja kahdessa partisiipimuodossa *состоящий из* (*состоящие из*) (koostuva, koostuvat). Koetinehdokas on käyttökelpoinen, sillä 10 validissa osumassa koostumussuhteen alakäsitteiden luettelo alkaa välittömästi koetinehdokkaan jälkeen ja yhdeksässä validissa osumassa koostumussuhteen yläkäsite on välittömästi ennen koetinehdokasta:

(221) Осушительная система состоит из насосов, трубопроводов, клапанов, приемных сеток, устройств сигнализации и замера уровня воды. (Bezop: 49)

Esimerkissä 221 yläkäsitteen nimitys *осушительная система* (kuivatusjärjestelmä) on välittömästi ennen koetinehdokasta ja alakäsitteiden luettelo *насосы* (pumput), *трубопроводы* (putkistot), *клапаны* (venttiilit), *приемные сетки* (suodattimet), *устройства сигнализации* (hälytyslaite) ja *замер уровня воды* (vedenpinnan mittauslaite) välittömästi koetinehdokkaan jälkeen. Useimmiten välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen onkin termi. Vain kolmessa validissa osumassa yksikään merenkulun turvallisuustermi ei ole välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen, vaan välissä on 1–2 sanaa. Näistä kahdessa osumassa aiemmin tekstissä mainittu yläkäsitteen nimitys on koettimen yhteydessä korvattu pronomiinilla:

(222) [Облегченный пластырь] - - Он имеет размеры 3х3м, состоит из двух полотнищ парусины водоупорной пропитки, положенных крест-накрест основной ткани относительно друг друга, и прокладки из грубошерстного войлока между ними. (Bezop: 52)

Esimerkissä 222 yläkäsitteen nimitys on korvattu pronomiinilla *он* (se). Yläkäsitteen nimitys *облегченный пластырь* (kevytвуотоматто) on mainittu edeltävän virkkeen alussa.

**Koetinehdokas *включать в себя*.** Manuaalisen haun tuottama koetinehdokas *включать в себя* (sisältää) ei ollut tämän tutkimuksen aineistossa aivan yhtä produktiivinen kuin koetinehdokas *состоять из*, mutta kaikki hakuehdon *включает в себя* aineistosta Bezop tuottamat osumat ovat valideja. Konkordanssihakua tuotti kuusi osumaa, joissa kaikissa koetinehdokas toimii koostumussuhteen koettimena. Kaikissa osumissa yläkäsitteen nimitys on ennen koetinehdokasta, kuten esimerkissä 223 termi *стояночная*

*операция судна* (aluksen satamatoiminnot). Alakäsitteiden luettelo on kaikissa osumissa koetinedokkaan jälkeen:

- (223) - - стояночная операция судна включает в себя грузовые операции, крепление груза и выполнение целого ряда вспомогательных операций; - - (Bezop: 36)

Yhdessä osumassa kerrotaan osasta, joka EI kuulu yläkäsitteen kokonaisuuteen:

- (224) Существующая система и практика отчетности не включает в себя и такого важного обстоятельства, как учет отказов систем и оборудования судна, - - (Bezop: 43)

Esimerkissä 224 koetinedokkaan ja alakäsitteen *учет отказов систем и оборудования судна* (hälytysloki) välissä on poikkeuksellisesti sanoja, kun taas kaikissa muissa osumissa alakäsitteiden luettelo seuraa välittömästi koettimen jälkeen. Mielenkiintoista on havaita, että neljässä osumassa käsitteet ovat toimintakäsitteitä ja vain kahdessa osumassa oliokäsitteitä.

*Eskos*-aineistosta poimittujen termien kontekstien tarkastelu tuotti viisi mahdollista koostumussuhteen substantiivikoetinta. Koetinedokkaat *элемент*, *компонент* ja *часть* viittaavat kokonaisuuden osiin kun taas *состав* ja *структура* viittaavat kokonaisuuteen. Aiemmissä tutkimuksissa on todettu, että myös kokonaisuutta tarkoittavat substantiivit *система* ja *комплекс* voivat toimia koostumussuhteen koettimina. Näiden mahdollisten substantiivikoettimien toimintaa testasin tekemällä aineistoista *Eskos* ja *Bezop* konkordanssihaut hakuehdoilla *элемент\**, *компонент\**, *часть\**, *в состав\**, *структур\**, *систем\** ja *комплекс\**.

**Koetinedokas *элемент*.** Konkordanssihaku artikkelista *Eskos* substantiivista *элемент* (osa) muodostetulla hakuehdolla *элемент\** tuotti 11 osumaa, joista kolmessa koetinedokas toimii koostumussuhteen koettimena. Näissä osumissa lauseen syntaktinen rakenne on muotoa *одним из элементов \* является*, jossa katkaisumerkin kohdalla on yläkäsitteen nimitys, kuten esimerkiksi 225, tai muotoa *является элементом*, jossa alakäsitteen tai alakäsitteiden nimitykset ovat ennen koetinedokasta ja yläkäsitteen nimitys koetinedokkaan jälkeen, kuten esimerkiksi 226:

- (225) Одним из элементов СУДС является информационная подсистема, аккумулирующая информацию в зоне действия СУДС о позициях судов в портах и на подходах к ним, перемещениях внутри порта и судоходах. (Eskos: 15)

- (226) - - CCC или АИС и является элементом информационной бассейновой ЕСКУС. (Eskos: 18)

Esimerkissä 226 jompikumpi alakäsitteistä *CCC* (alusilmoitusjärjestelmä) tai *АИС* (AIS) on osa yläkäsitteen *информационная бассейновая ЕСКУС* (informatiivinen alueellinen VTMS) kokonaisuutta. Muut kuin validit osumat ovat esimerkkejä käsitteestä, joka on kokonaisuus, kuten ЕСКУС esimerkissä 227, mutta jonka koostumussuhteisia alakäsitteitä ei lähikontekstissa ole annettu eksplisiittisesti:

- (227) Для обмена информацией с элементами ЕСКУС в заливе Петра Великого в распоряжение ТОИТЦ могут быть предоставлены высокоскоростные выделенные каналы связи, создаваемые по проекту ЕСПВ. (Eskos: 19)

Aineistosta *Bezop* hakuehto *элемент\** tuotti 46 osumaa. Niistä vain seitsemän on sellaisia, joissa koetin toimii koostumussuhteen koettimena:



- (228) Основными элементами системы безопасности мореплавания являются: человек, судно как инженерное сооружение, груз как объект перевозки, и морская стихия – среда, в которой морской транспорт осуществляет свою деятельность. (Bezop: 39)

Esimerkissä 228 *система безопасности мореплавания* (merenkulun turvallisuusjärjestelmä) on koostumussuhteen yläkäsite ja *человек* (ihminen), *судно* (alus), *груз* (lasti) ja *морская стихия* (meriympäristö) alakäsitteitä. Lauseen syntaktisesta rakenteesta johtuen koetinedokas *элементами* on tekstissä sekä yläkäsitteen että alakäsitteiden nimitysten edellä. Tässä esimerkissä koetinedokkaaseen liittyy verbimuoto *являются*. Jokin verbin *являться* muoto on valideista seitsemästä osumasta neljässä. Verbimuotoa ei kuitenkaan voi liittää hakuehtoon suoraan substantiivin yhteyteen, koska verbin ja koettimen välissä saattaa olla monta sanaa.

- (229) Груз как элемент транспортной системы оказывает большое влияние на безопасность плавания (Bezop: 17)

Esimerkissä 229 koostumussuhteinen yläkäsite on *транспортная система* (kuljetusjärjestelmä) ja alakäsite *груз* (lasti). Seitsemästä validista osumasta kahdessa on appositio *как элемент*, kuten esimerkissä 229.

Seitsemän validin osuman lisäksi konkordanssilistassa on 20 sellaista osumaa, josta ilmenee, että käsite on kokonaisuus, jolla on osia, mutta osakäsitteitä ei mainita ainaakaan koetinedokkaan lähikontekstissa:

- (230) Для этого необходимо исследовать поэтапно безопасность отдельных элементов всех систем судна и судно как систему в целом с учетом его физического износа и морального старения оборудования. (Bezop: 43)

Esimerkissä 230 ilmenee, että laiva on järjestelmä, joka koostuu järjestelmistä, jotka puolestaan koostuvat osista, mutta esimerkissä ei mainita, mitkä nämä osat ovat. Joskus kuitenkin pelkkä tieto siitä, että käsite on osista koostuva kokonaisuus, voi olla tarpeellinen.

**Koetinedokas *компонент*.** Substantiivista *компонент* (osa) muodostetulla hakuehdolla *компонент\** tehty konkordanssihaku *Eskos*-artikkelista tuotti kolme osumaa, joista jokaisessa koetin toimii koostumussuhteen koettimena:

- (231) Одним из компонентов АИС является радиомаяк DGPS - - (Eskos: 16)

Esimerkissä 231 sekä yläkäsite *АИС* (AIS) että alakäsite *радиомаяк DGPS* (DGPS-maa-asema) ovat tekstissä koetinedokkaan *компонент* jälkeen. Koetinedokkaaseen liittyy tässä esimerkissä verbimuoto *является*. Substantiivi *компонент* näyttäisi siis käyttäytyvän samoin kuin substantiivi *элемент*, sillä kummankin tuottamissa osumissa on usein samanlainen syntaktinen rakenne, jossa substantiivia seuraa verbi *являться*.

*Bezop*-aineistosta tehty konkordanssihaku hakuehdolla *компонент\** tuotti vain yhden osuman, joka ei kuitenkaan ole kelvollinen esimerkki koettimen toiminnasta. Kun verrataan aineistojen laajuutta sanojen lukumäärällä mitattuna, aineistossa *Bezop* on sanoja noin 13 kertaa niin paljon kuin artikkelissa *Eskos*. Jos aineiston *Bezop* kirjoittajat olisivat käyttäneet sanaa *компонент* keskimäärin yhtä usein kuin artikkelin *Eskos* kirjoittajat, hakuehdon olisi pitänyt tuottaa noin 40 osumaa. Sanan erilainen frekvenssi

teksteissä osoittaa, että koettimet ovat kirjoittajakohtaisia. Kumpaankin tekstiin pätee kuitenkin, että sana *элемент* on yleisempi kuin sana *компонент*.

**Koetinehdokas *часть*.** Konkordanssihaku artikkelista *Eskos* substantiivin *часть* (osa) hakuehdolla *часть*\* tuotti neljä osumaa, joista yhdessä koetin toimi koostumussuhteen koettimena. Tosin sekä kokonaisuus *ЕСКУС Тихоокеанского бассейна* että sen osa *ТОИТЦ* ovat yksilökäsitteitä eivätkä näin ollen terminologisesti kovin kiinnostavia:

- (232) ТОИТЦ становится центральной частью ЕСКУС Тихоокеанского бассейна, - -  
(Eskos: 19)

Aineistosta *Bezop* hakuehto *часть*\* tuotti 17 osumaa, joista neljässä koetinehdokas toimii koostumussuhteen koettimena. Neljästä validista osumasta kolmessa substantiiviin *часть* liittyy adjektiivimäärite *составная*:

- (233) Комплекс таких действий является составной частью борьбы за живучесть судна - -  
(Bezop: 70)

Esimerkissä 233 tiettyjen toimenpiteiden kokonaisuus on yläkäsitteen *борьба за живучесть судна* (aluksen pelastaminen) osa, mutta alakäsitteitä eli toimenpiteitä on etsittävä konkordanssihaun tuottaman kontekstin ulkopuolelta. Samoin kuin substantiiveihin *элемент* ja *компонент* myös koetinehdokkaaseen *часть* liittyy usein verbi *является*, mutta syntaktinen rakenne eroaa siten, että nyt verbi on ennen substantiivia. Valideista neljästä osumasta kolmessa lauseen syntaktinen rakenne on *что-то является составной частью чего-то*.

**Koetinehdokas *состав*.** Kokonaisuuteen viittaavista koetinehdokkaista tarkastelen ensin substantiivia *состав* (kokoonpano, koostumus), josta muodostin hakuehdon *в состав*\*. Artikkelista *Eskos* hakuehto tuotti kaksi osumaa, joissa kummassakin koetinehdokas toimii koostumussuhteen koettimena:

- (234) Следовательно, ТОИТЦ может одновременно действовать в качестве центра региональной информационной системы в составе ЕСКУС залива Петра Великого.  
(Eskos: 19)

Esimerkissä 234 koetinehdokas *в составе* osoittaa, että alakäsite *центр региональной информационной системы* (alueellinen informaatiopalvelukeskus) on koostumussuhteen yläkäsitteen *ЕСКУС* (VTMIS) osa. Tässä esimerkissä alakäsite on välittömästi ennen koetinehdokasta ja yläkäsite välittömästi sen jälkeen.

Aineistosta *Bezop* hakuehto *в состав*\* tuotti kahdeksan osumaa, joista kolmessa koetinehdokas toimii koostumussuhteen koettimena:

- (235) Цемент, как известно, входит в состав аварийного снабжения и имеется на каждом судне в количестве от 100 до 400 кг в зависимости от типа и размера судна. (Bezop: 75)

Esimerkissä 235 alakäsite *цемент* (sementti) on ennen koetinehdokasta *в состав* ja yläkäsite *аварийное снабжение* (vaurion torjuntamateriaali) välittömästi koetinehdokkaan jälkeen. Hakuehto tuotti aineistoista *Eskos* ja *Bezop* yhteensä viisi validia osumaa. Niistä kolmessa koetinehdokkaaseen liittyy verbin *входить* yksikön 3. persoonan preesensmuoto *входит* tai monikon 3. persoonan preesensmuoto *входят*. Lauseen syntakti-

nen rakenne on siis *что-то входит в состав чего-то*. Hakuehtoa voi tarkentaa lisäämällä siihen verbin *входить* (*входит в состав/входят в состав*), jolloin roskan määrä vähenee. Keino on kuitenkin käyttökelpoinen vain suurella aineistolla, koska tarkennetulla hakuehdolla menetetään valideja osumia.

**Koetinehdokas *структура*.** Toisella kokonaisuuteen viittaavalla koetinehdokkaalla, substantiivilla *структура* (rakenne) aineistosta *Eskos* tehty konkordanssihaku tuotti hakuehdolla *структур\** seitsemän osumaa, joista kolmessa koetinehdokas toimii koostumussuhteen koettimena:

- (236) - - включение в структуру ЕСКУС четвертой технической составляющей – информационной подсистемы, - - (Eskos: 17)

Esimerkissä 236 koostumussuhteen yläkäsite on *ЕСКУС* (VTMIS) ja yksi sen alakäsitteistä on *информационная подсистема* (informaatiopalvelu). Aineistosta *Bezop* hakuehto tuotti seitsemän osumaa, joista yksikään ei varsinaisesti ole esimerkki toiminnasta koostumussuhteen koettimena. Kahdessa osumassa kuitenkin kuvataan järjestelmän rakennetta:

- (237) Система «человек – машина» имеет иерархическую структуру, на вершине которой находится человек, - - (Bezop: 13)

Esimerkissä 237 sanotaan, että käsitteellä *система «человек – машина»* (ihminen-kone-järjestelmä) on hierarkkinen rakenne ja että hierarkian huipulla on ihminen. Koetinehdokas näyttäisi kuitenkin lähinnä ilmaisevan, että käsitteellä on jonkinlainen rakenne, mutta koostumussuhteen ylä- ja alakäsitteiden ilmaisuun koetinehdokas ei tämän tutkimuksen aineistossa ole erikoistunut.

**Koetinehdokas *эман*.** Manuaalinen koetinhaku aineistosta *Bezop* tuotti yhden koostumussuhteen mahdollisen koettimen, substantiivin *эман* (vaihe). Koetinehdokasta *эман* voidaan pitää myös aikasuhteen koettimena, koska se toimii ajallisesti toisiaan seuraavien tapahtumien tai toimintojen koettimena. Tarkastelen sitä kuitenkin mahdollisena koostumussuhteen koettimena, koska käyttämässäni käsitesuhteiden luokittelussa prosessi on kokonaisuus, joka koostuu vaiheista. Tässä luokittelussa vaiheet ovat prosessin koostumussuhteisia alakäsitteitä. Konkordanssihaku hakuehdolla *эман\** aineistosta *Bezop* tuotti 12 osumaa, joista neljässä on sekä koostumussuhteen ylä- että alakäsite:

- (238) Так, например, первая – движеническая операция судна может быть разбита на три этапа: управление судном в открытом море, управление судном в зонах разделения движения судов и на подходах к портам, управление судном в закрытых портовых акваториях; - - (Bezop: 36)

Esimerkissä 238 on ennen koetinehdokasta mainittu koostumussuhteen yläkäsite *движеническая операция судна* (aluksen ohjailu), joka jakautuu kolmeen vaiheeseen: aluksen ohjailuun avomerellä, aluksen ohjailuun liikenteenjakoalueella ja sisääntuloväylillä sekä aluksen ohjailuun satama-altaassa. Neljän validin osuman lisäksi hakuehto tuotti kolme osumaa, joissa on vain koostumussuhteen yläkäsite ja kaksi osumaa, joissa on vain alakäsite.

**Koetinehdokkaat *система* ja *комплекс*.** Tunnettujen termien kontekstien tarkastelu *Eskos*-artikkelista ja manuaalinen koostumussuhteen koetinhaku *Bezop*-aineistosta eivät

tuottaneet yhtään kontekstia, jossa kokonaisuuteen viittaavat substantiivit *система* (järjestelmä) ja *комплекс* (kokonaisuus) toimisivat koostumussuhteen koettimina. Koska nämä substantiivit kuitenkin ovat mahdollisia koettimia aiempien tutkimusten perusteella, tein konkordanssihaut kummastakin aineistosta myös hakuehdoilla *систем\** ja *комплекс\**.

Hakuehto *систем\** tuotti *Eskos*-artikkelista 46 osumaa ja *Bezop*-aineistosta 254 osumaa. Osumien runsaslukaisuus artikkelissa selittyy osittain sillä, että substantiivi *система* on termin osana 27 osumassa:

- (239) Советом Службы морского флота Минтранса РФ в сентябре 1998 г. было принято Постановление N 6 о единой системе контроля и управления судоходством в прибрежных водах России, основное внимание в котором уделяется развитию АИС в России и созданию на основе АИС, СУДС и районов А1 ГМССБ единой системы контроля и управления судоходством. (*Eskos*: 16)

Esimerkissä 239 *система* on osa termiä *единая система контроля и управления судоходством* (alusliikenteen ohjaus- ja informaatiopalvelut). Sana *система* on usein termin osa todennäköisesti myös *Bezop*-aineistossa, vaikka tapauksia ei ole tässä laskettu, koska *Bezop*-aineistosta ei termejä ole tunnistettu. Esimerkissä 240 tämä substantiivi on osa sanaliittoja *система водяных завес* (vesiverhojärjestelmä) ja *осушительная система* (kuivatusjärjestelmä), jotka ovat mielestäni merenkulun turvallisuuden termejä:

- (240) Если известно место очага пожара в трюме, можно отсечь его от остального груза, запустив систему водяных завес в этом месте одновременно с осушительной системой. (*Bezop*: 115)

*Bezop*-aineistosta hakuehto tuotti 13 sellaista osumaa, joissa on koostumussuhteen ylä- ja alakäsite. Substantiivi *система* on kaikissa näissä osumissa yläkäsitteen nimityksen osana, ja suurimmassa osassa osumista koostumussuhdetta osoittaa jompikumpi koetinehdokkaista *состоять из* (koostua) tai *элемент* (osa), kuten esimerkiksi 241:

- (241) Основными элементами системы безопасности мореплавания являются: человек, судно как инженерное сооружение, груз как объект перевозки, и морская стихия – среда, в которой морской транспорт осуществляет свою деятельность. (*Bezop*: 39)

Muissa osumissa substantiivilla on usein yleiskielen merkitys 'kokonaisuus':

- (242) В зависимости от особенностей применяемой аппаратуры АИС может непосредственно интегрироваться СУДС, создавая единую базу оперативных данных, или же образует независимую систему с отдельной базой данных. (*Eskos*: 18)

*Система* ei siis tämän tutkimuksen aineistossa toimi koostumussuhteen koettimena, koska se ei tuota sellaisia osumia, joissa on sekä koostumussuhteen ylä- että alakäsite. Koetinehdokas antaa kuitenkin tietoa käsitteiden laadusta. Jos sana liittyy käsitteen nimitykseen, se osoittaa, että käsite on kokonaisuus ja kokonaisuus voidaan aina jakaa osiin.

Substantiivi *комплекс* ei toimi koettimena tutkimukseni aineistossa, sillä hakuehto *комплекс\** ei tuottanut yhtään validia osumaa. Substantiivilla on varsin alhainen frekvenssi tutkimukseni aineistosta, sillä aineistosta *Eskos* hakuehto ei tuottanut yhtään osumaa ja laajemmasta *Bezop*-aineistostakin se tuotti vain 12 osumaa.

**Yhteenveto.** Tutkimusaineistossani vahvimaksi venäläiseksi koostumussuhteen koetimiksi osoittautui verbi *состоять из*, joka on melko produktiivinen ja tuottaa vähän roskaa. Myös verbi *включать в себя* on kelvollinen koostumussuhteen koetin. Se ei ollut yhtä produktiivinen kuin *состоять из* mutta ei tuottanut yhtään roskaa. Kumpikin verbikoetin esiintyy tekstissä yhdessä termin kanssa, ja usein niitä seuraa alakäsitteiden luettelo. Substantiivikoettimista vähiten roskaa tuotti *компонент*. Se saattaa kuitenkin olla kirjoittajakohtainen. Substantiivin *состав* tuottamista osumista joka toinen on validi, mutta se ei ole yhtä produktiivinen kuin verbikoettimet. Laajasta aineistosta tehdyssä konkordanssihaussa roskan määrää voi vähentää entisestään tarkentamalla hakuehto lisäämällä siihen verbin *входить* (*входит в состав/входят в состав*), jolloin tosin menetetään joitakin valideja tapauksia. Muut substantiivikoettimet tuottivat enemmän roskaosumia kuin valideja koostumussuhdetapauksia. Lisäksi ne saattavat olla kirjoittajakohtaisia. Hakuehto *элемент\** tuotti validien osumien lisäksi osumia, joissa on koostumussuhteen yläkäsite, mutta sen alakäsitteitä ei lähikontekstissa ole annettu eksplisiitisti.

### 7.5.6 Vieruskäsitteiden venäläiset koettimet

Artikkelilla *Eskos* tehty koettimien poimintakoe tuotti mahdollisiksi vieruskäsitteitä osoittaviksi venäläisiksi koettimiksi sulkeet, sanaliiton *в том числе* (muun muassa) sekä konjunktioit *и* (ja) ja *или* (tai). Myös sanaliitto *и прочие* (ja muut), lyhenteet *и т.д.* (jne.) ja *и др.* (ym.) sekä jälkimmäisen lyhenteen pitkä muoto *и другое* (ynnä muut) ovat manuaalisen haun perusteella mahdollisia vieruskäsitteiden koettimia. Tärkeimmät konkordanssihaun tulokset olen koonnut taulukkoon 53.

TAULUKKO 53. Vieruskäsitteiden terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
sulkeet	<i>(*</i>	370	33	9	heikko
и др.	<i>и др.</i>	16	5	31	heikko
в том числе	<i>в том числе</i>	7	2	29	heikko

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *sulkeet*.** Olen aiemmin todennut, että sulkeet toimivat eräänlaisena yleiskoettimena, joka ilmaisee monenlaista terminologisesti kiinnostavaa tietoa, kuten esimerkiksi vieruskäsitesuhdetta. Hakuehdolla *(\** tehty konkordanssihaku aineistosta *Eskos* tuotti viisi sellaista tulosta, jossa sulkeissa on vähintään kaksi käsitettä luettelon muodossa. Koska käsitteet on ilmaistu luettelona, voidaan olettaa niiden liittyvän jollakin tavalla toisiinsa. Ne saattavat olla vieruskäsitteitä siinä käsitejärjestelmässä, joka tekstin avulla voidaan rakentaa. Esimerkiksi käsitteet *судовладелец* (laivanisäntä), *агент* (laiva-asiamies) ja *портовик* (ahtaaja) eivät ehkä ole vieruskäsitteitä tai edes lähikäsitteitä talous- tai koulutusalan käsitejärjestelmässä, mutta kuljetusalan käsitejärjestelmässä ne voivat hyvinkin olla vieruskäsitteitä:

(243) - - и для прочих участников транспортного процесса (судовладельцев, агентов, портовиков и т.д.). (Eskos: 18)

Esimerkki 243 on valittu tähän jaksoon kuvaamaan sulkeiden toimintaa vieruskäsitteiden koettimena. Samassa lauseessa on kuitenkin kaksi muutakin koetinehdokasta, jotka saattavat osoittaa vieruskäsitteiden luetteloa. Nämä koetinehdokkaat ovat sanaliitto *и прочие* (ja muut) ja lyhenne *и т.д.* (jne.). Lyhenne luettelon lopussa osoittaa, että luettelo ei ole täydellinen. Ennen sulkeita on geneerinen yläkäsite, jota edeltää adjektiivi *прочие*. Aineistosta *Bezop* tehdyn manuaalisen koetinhaun perusteella myös lyhenne *и др.* (ym.) ja sen pitkä muoto *и другие* (ynnä muut) ovat mahdollisia vieruskäsitteiden koettimia. Konkordanssihakua hakuehdolla *и др.* tuotti 16 osumaa, joista 12:ssa lyhennettä edeltää merenkulun turvallisuuteen liittyvien käsitteiden luettelo. Viidessä tapauksessa konkordanssilistaan tulostunut konteksti sisältää lueteltujen käsitteiden yläkäsitteen nimityksen, jolloin luettelon käsitteet ovat todennäköisesti vieruskäsitteitä tekstin käsitejärjestelmässä. Olen käsitellyt näitä koetinehdokkaita tarkemmin jaksossa 7.5.4 geneerisen käsitesuhteen koetinehdokkaana.

Aineistosta *Bezop* konkordanssihakua hakuehdolla (\* tuotti 29 sellaista tulosta, jossa sulkeissa on vähintään kahden käsitteen luettelo:

- (244) Порошковые составы применяют для тушения горящих металлов (калия, натрия, лития, магния). (Bezop: 81–82)
- (245) - - содержать рекомендации по некоторым, нигде не освещенным вопросам (затопление трюма, паника, сбрасывание каравана и др.). (Bezop: 64)
- (246) - - когда плотность груза меньше плотности забортной воды (нефтепродукты, спирты), то процесс замещения кончается, - - (Bezop: 100)

Esimerkissä 244 sulkeissa on luettelo metalleista ja välittömästi ennen sulkeita on mahdollisen yläkäsitteen nimitys *горящий металл* (palava metalli). Sulkeissa olevat metallit ovat aivan ilmeisesti vieruskäsitteitä ja niiden yhteinen yläkäsite on *горящий металл*. Osumaluettelossa on 23 vastaavaa tapausta, jossa välittömästi ennen sulkeita on yläkäsitteen nimitys ja sulkeissa on alakäsitteiden luettelo. Neljässä osumassa ennen sulkeita on käsite, joka ei voi olla sulkeissa olevien käsitteiden yläkäsite ainakaan merenkulun turvallisuuden kontekstissa. Esimerkissä 245 ennen sulkeita on yleiskielinen sana *вопрос* (kysymys). Kahdessa tapauksessa ennen sulkeita on annettu käsitepiirre, joka on yhteinen sulkeissa mainituille käsitteille. Esimerkissä 246 sulkeissa ovat käsitteet *нефтепродукты* (öljytuotteet) ja *спирты* (alkoholit), joiden yhteinen käsitepiirre on, että niiden tiheys (плотность) on meriveden tiheyttä pienempi.

Sekä *Eskos*- että *Bezop*-aineistojen osumalistojen perusteella voidaan sanoa, että sulkeet toimivat vieruskäsitteiden venäläisenä koettimena, mutta ne tuottavat erittäin runsaasti kelpaamattomia osumia. Sulkeissa olevien käsitteiden suhteen selvittämiseksi on tarkasteltava ympäröivää kontekstia. Jos välittömästi ennen sulkeita on mainittu käsite, joka on sulkeissa annettujen käsitteiden yläkäsite tekstin käsitejärjestelmässä, ovat sulkeissa annetut käsitteet todennäköisesti vieruskäsitteitä tässä käsitejärjestelmässä.

**Koetinehdokas в том числе.** Sulkeiden lisäksi koettimien poiminta artikkelista *Eskos* tuotti sanaliiton *в том числе* (muun muassa, mukaan lukien), joka on mahdollinen vieruskäsitesuhteen koetin. Hakuehto *в том числе* artikkelista *Eskos* tuotti neljä osumaa, joista kaksi on validia. Näissä osumissa sanaliittoa seuraa vieruskäsitteiden luettelo:



- (247) Второй уровень образуют региональные ЕСКУС, охватывающие акватории нескольких портов, в том числе подходные пути и зоны разделения движения. (Eskos: 18)

Esimerkissä 247 koetinehdokas *в том числе* (muun muassa, mukaan lukien) osoittaa, että käsitteet *подходный путь* (tuloväylä) ja *зона разделения движения* (liikenteenjakovyöhyke) ovat lähikäsitteitä, jotka liittyvät käsitteeseen *акватория порта* (satama-alueen vesialue). Koska esimerkissä puhutaan käsitteistä, joilla on sijaintipiirre, looginen tulkinta on, että käsitteillä on lokaalinen yhteenkuuluvuus. Niiden voidaan ajatella olevan yläkäsitteen *региональная ЕСКУС* (alueellinen VTMS) koostumussuhteisia alakäsitteitä, jolloin ne ovat keskenään vieruskäsitteitä.

Aineistosta *Bezop* konkordanssihaku hakuehdolla *в том числе* tuotti kolme osumaa, joista yhdessä ei sanaliittoon liittynyt käsitteiden luetteloa tai käsitettä. Tämä osoittaa, että koettimet ovat hyvin teksti- ja kirjoittajakohtaisia.

**Koetinehdokkaat *и* ja *или*.** Sulkeiden ja sanaliiton *в том числе* lisäksi mahdollisia vieruskäsitteiden koettimia ovat konjunktiot *и* (ja) ja *или* (tai). Konkordanssihaku artikkelista *Eskos* hakuehdolla *и* tuotti 147 osumaa, joista neljässä konjunktio sitoo yhteen kahta termiä. Kaikissa näissä osumissa toinen termeistä on elliptinen:

- (248) Дополнительная информация о судах, получаемая с помощью АИС (тип судна, род и количество груза, агент или судовладелец, порты отправления и назначения) может эффективно использоваться в информационных подсистемах СКЦ (СПЦ) и СУДС. (Eskos: 16)

Esimerkissä 248 on termiparit *под груза* (tavaralaji) ja *количество груза* (tavaramäärä) sekä *порт отправления* (lähtösatama) ja *порт назначения* (määräsatama). Termien yhteiset osat *груз* ja *порт* ovat tekstissä kuitenkin vain kerran.

Hakuehto *или* tuotti 19 osumaa. Niistä kahdessa konjunktio on sitomassa yhteen kahta termiä, joilla on yhteinen osa, toisin sanoen toinen termeistä on elliptinen:

- (249) - - реальностью создания отдельных районов А2 или А1 при небольших портах, - - (Eskos: 19)

Esimerkissä 249 konjunktio *или* rinnastaa termit *район А2* (alue A2) ja *район А1* (alue A1), joiden yhteinen osa *район* on tekstissä vain kerran. Kolmessa osumassa kahta termiä sitoo yhteen konjunktioiden yhteenliittymä *и/или*, kuten esimerkissä 250, jossa se rinnastaa käsitteet *радиолокационный контроль* (tutkavalvonta) ja *транспондерный контроль* (AIS-valvonta):

- (250) Европы непрерывной зоной радиолокационного и/или транспондерного контроля за движением судов; - - (Eskos: 16)

Aineistosta *Bezop* haku tuotti vastaavasti 1 340 ja 232 osumaa. Näin suuri osumamäärä osoittaa, etteivät konjunktiot *и* ja *или* voi sellaisinaan toimia koettimina. En käsittele tässä konkordanssihaun tuloksia hakuehdoilla *и* ja *или* aineistosta *Bezop*, koska haku tuotti niin runsaasti tuloksia, ettei niiden manuaalinen käsittely ole mielekästä.

**Yhteen veto.** Tutkimukseni ei tuottanut yhtään vahvaa tai hyvää vieruskäsitteiden venäläistä koetinta. Koetinehdokkaista produktiivisin on sulkeet, mutta niiden tarkkuus on erittäin alhainen. Tarkkuus paranee huomattavasti, jos vieruskäsitteiden yläkäsite on

tunnettu, jolloin yläkäsitteen nimityksen voi lisätä hakuehtoon sulkeiden eteen. Tarkimmaksi koetinehdokkaaksi osoittautui lyhenne *u op.*, mutta senkin tuottamista osuista vain kolmannes on valideja. Konjunktiot *u* ja *ulu* liittyvät elliptisiin ilmauksiin, mutta koettimina ne toimivat vain yhdessä termin kanssa.

### 7.5.7 Funktiosuhteiden venäläiset koettimet

Tunnetuilla termeillä tehty koetinhaku aineistosta *Eskos* ei tuottanut yhtään kontekstia, jossa käsitteiden välillä olisi kausaalisuhde, vuorovaikutussuhde tai toimintasuhde. Käsitteiden välisistä funktiosuhteista aineistossa on esimerkkejä vain spatiaalisesta funktiosuhteesta, jossa käsitteiden välillä on tilasuhde. Manuaalinen koetinhaku aineistosta *Bezop* tuotti kuitenkin konteksteja, joissa käsitteiden välillä on muita funktiosuhteita. Noudatan funktiosuhteisiin liittyvien venäläisten koettimien tarkastelussa samanlaista ryhmittelyä kuin suomalaisten koettimien tarkastelussa (ks. jakso 7.4.7) ja aloitan kausaalisuhteen koettimista.

#### 7.5.7.1 Kausaalisuhteiden venäläiset koettimet

Syy- ja seuraussuhteessa syy voi olla toiminta tai tapahtuma, mutta se voi olla myös olio, tila tai olosuhde. Seuraus voi sekin olla esimerkiksi tila (ks. jakso 7.4.7.1). Siksi onkin yllättävää, että konkordanssihakua aineistosta *Eskos* ei tuottanut yhtään syytä tai seurausta osoittavaa koetinehdokasta. Tarkoituksen mahdollisia koettimia haku tuotti yhden, verbin *обеспечивать* (turvata, taata), mutta kaikissa osumissa puhutaan toiminnan tarkoituksen sijasta oliokäsitteen tarkoituksesta. Tästä syystä en tarkastele verbiä *обеспечивать* tässä yhteydessä. Sen sijaan aineistosta *Bezop* manuaalinen koetinhaku tuotti runsaasti mahdollisia syy- ja seuraussuhteen sekä tarkoituksen koettimia. Koetinehdokkaiden joukosta valitsin ne, jotka tuottivat manuaalisessa haussa vähintään kolme kontekstia, ja tein konkordanssihaut mahdollisilla verbikoettimilla *вызвать* (aiheuttaa) sekä *приводить* ja *привести* (johtaa), mahdollisilla substantiivikoettimilla *результат* (tulos), *причина* (syy) ja *вследствие* (seuraus), *цель* (tarkoitus, tavoite), adverbilla *следовательно* (niin muodoin, siis) ja prepositiolla *из-за*. Konkordanssihakujen tulokset olen koonnut taulukkoon 54.

TAULUKKO 54. Kausaalisuhteen terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
привести	<i>привести/приведет/привело</i>	22	17	77	vahva
приводить	<i>приводит/приводят</i>	15	12	80	vahva
вызвать, вызвать	<i>вызвать/вызывает</i>	14	13	93	vahva
результат	<i>в результате</i>	42	26	62	vahva
причина	<i>причин*</i>	49	11	22	heikko
следствие	<i>*следстви*</i>	26	10	38	hyvä
цель	<i>целью/в целях</i>	17	6	35	hyvä
следовательно	<i>следовательно</i>	26	9	35	hyvä
из-за	<i>из-за</i>	13	8	62	vahva

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *привести*.** Manuaalisessa koetinhaussa mahdollisista syyn ja seurauksen verbikoettimista produktiivisin oli verbi *привести* (johtaa), joka voi koettimena toimiessaan olla perusmuodon lisäksi yksikön kolmannen persoonan preesensmuodossa *приведет* tai preteritimuodossa *привело*. Manuaalisen koetinhaun perusteella tein konkordanssihaun aineistosta *Bezop* hakuehdolla *привести/приведет/привело*. Haku tuotti 22 osumaa, joista 17 osumassa verbi toimii merenkulun turvallisuuskäsitteiden välisen syyn ja seurauksen koettimena. Kymmenessä validissa osumassa verbi on perusmuodossa, kolmessa osumassa muodossa *приведет* ja neljässä preteritimuodossa *привело*:

(251) Пожар и взрывы в грузовом трюме могут отразиться на эффективности крепления груза, что в условиях качки приведет его смещению и потере остойчивости. (Bezop: 115)

(252) - - в трюме произошло продавливание обшивки танка, что привело к затоплению трюма, - - (Bezop: 17)

Esimerkissä 251 palo (пожар) ja räjähdykset (взрывы) voivat välillisesti olla lastin siirtymisen (смещение груза) syy. Esimerkissä 252 tankin sauman repeäminen (продавливание обшивки танка) on aiheuttanut ruuman kastumisen (затопление трюма). Kummassakin esimerkissä seurauskäsitteen nimitys on välittömästi koetinehdokkaan jälkeen. Kolmea tapausta lukuun ottamatta kaikissa valideissa tapauksissa koetinehdokkaaseen liittyy välittömästi prepositio *к*, kuten esimerkissä 252.

Verbin toimiessa syy- ja seuraussuhteen koettimena siihen liittyy joka toisessa tapauksessa apuverbin yksikön kolmannen persoonan muoto *может* (voi) tai monikon kolmannen persoonan muoto *могут* (voivat). Syykäsite ei siis välttämättä johda seurauskäsitteeseen:

(253) Пар портит ценные грузы, вызывает разбухание зерна и волокнистых грузов, что может привести к деформации корпуса судна. (Bezop: 80–81)

Esimerkissä 253 on viitattu kahteen syykäsitteeseen, jotka ovat *viljalastin turpoaminen* (разбухание зерна) ja *kuitumaisen lastin turpoaminen* (разбухание волокнистых грузов), ja yhteen mahdolliseen seurauskäsitteeseen, joka on *rungon muodonmuutos* (де-

формации корпуса). Kaikissa valideissa tapauksissa verbiin liittyy prepositio *к*. Preposition voi lisätä hakuehtoon, koska se on kolmea poikkeusta lukuun ottamatta tekstissä välittömästi verbin jälkeen. Prepositiosta ei kuitenkaan ole hyötyä osumien ryhmittelyssä kelvollisiin ja kelpaamattomiin, sillä prepositio liittyy verbiin lähes kaikissa osumissa. Kelpaamattomissakin osumissa on käsitteiden välinen syy- ja seuraussuhde, mutta niissä syy ja seuraus eivät ole merenkulun turvallisuuden käsitteitä.

**Koetinehdokas *приводить*.** Verbikoettimista toiseksi produktiivisin oli manuaalisessa koetinhaussa verbin *привести* imperfektiivinen aspektipari *приводить* (johtaa), joka on koettimena toimiessaan joko yksikön 3. persoonan preesensmuodossa *приводит* tai monikon 3. persoonan preesensmuodossa *приводят*. Konkordanssihaku hakuehdolla *приводит/приводят* tuotti 15 osumaa, joista 12 on kelvollisia tapauksia koetinehdokkaan toiminnasta syyn ja seurauksen koettimena:

- (254) Применение углекислоты над сильно нагретой поверхностью приводит к образованию СО, который горюч. (Bezop: 120–121)
- (255) Например, столкновение приводит к пробоине, что вызывает разрушение связей и ослабление общей продольной прочности, затопление части внутреннего объема, разлив нефти, повышенную загазованность, искрообразование, взрыв или пожар. (Bezop: 101)

Esimerkissä 254 hiilidioksidin käyttö (применение углекислоты) on syy, jonka seurauks on hiilimonoksidin muodostuminen (образование СО). Esimerkkivirkkeen konteksti on itsestään syttyneen kivihiililastin sammuttaminen. Seurauskäsitteen nimitys on välittömästi koetinehdokkaan jälkeen, mutta syykäsitteen nimityksen ja koetinehdokkaan välissä on kuvattu olosuhde, jossa syyn ja seurauksen välinen suhde toteutuu. Palavaa hiilimonoksidia muodostuu, jos kuumen kivihiililastin pinnalle päästetään hiilidioksidia. Esimerkissä 255 yhteentörmäys (столкновение) on syy, jolla on seurausten ketju. Ensimmäisenä ketjussa on repeämä (пробоина), jolla puolestaan on monta vaihtoehtoista seurausta. Suurimmassa osassa valideja osumia onkin mainittu enemmän kuin yksi syy ja yksi seuraus. Verbin yksikkömuoto oli huomattavasti monikkomuotoa produktiivisempi, sillä monikkomuoto tuotti vain kaksi osumaa, joista yksi on kelvollinen. Samoin kuin verbiin *привести* myös verbiin *приводить* liittyy kaikissa valideissa tapauksissa prepositio *к*. Sen voi lisätä hakuehtoon, sillä se on välittömästi verbin jälkeen kaikissa muissa paitsi yhdessä validissa tapauksessa.

**Koetinehdokkaat *вызвать* ja *вызывать*.** Manuaalisen haun perusteella muotoiltu hakuehto *вызвать/вызывает* tuotti 14 osumaa, joista 13 on valideja. Seitsemässä osumassa on verbin perfekttiivisen aspektin infinitiivi *вызвать* (aiheuttaa) ja modaaliapuverbin *мочь* muoto *может* tai *могут*. Nämä osumat ovat kaikki valideja tapauksia koetinehdokkaan toiminnasta syyn ja seurauksen koettimena:

- (256) Остановка вентиляции может вызвать падение давления в МО - - (Bezop: 78)

Esimerkissä 256 syykäsitteen nimitys *остановка вентиляции* (ilmastoinnin pysähtyminen) on välittömästi ennen koetinehdokasta ja siihen liittyvää apuverbiä ja seurauskäsitteen nimitys *падение давления* (paineen aleneminen) välittömästi koetinehdokkaan jälkeen. Imperfekttiivisen aspektin yksikön 3. persoonan preesensmuoto *вызывает* (aiheuttaa) tuotti seitsemän osumaa, joista kuudessa koetinehdokas toimii syyn ja seurauk-

sen koettimena. Seuraus on näissä tapauksissa välittömästi koetinehdokkaan jälkeen, mutta syy saattaa olla tekstissä paljon ennen koetinta:

- (257) Пар портит ценные грузы, вызывает разбухание зерна и волокнистых грузов, что может привести к деформации корпуса судна. (Bezop: 80–81)
- (258) Высокократная пена в отличие от других не портит груз, отделку помещений и не вызывает коррозии, неопасна для здоровья человека, неэлектропроводна. (Bezop: 59)

Esimerkeissä 257 ja 258 seurauskäsitteen nimitykset *разбухание* (turpoaminen) ja *коррозия* (korroosio) ovat välittömästi koetinehdokkaan jälkeen, mutta syykäsitteen nimitys *пар* (höyry) on kolmen sanan ja *высокократная пена* (kevytvaaho) 11 sanan päässä ennen koetinehdokasta. Esimerkissä 258 ei tarkkaan ottaen ole syytä ja seurausta vaan käsitteet, jotka eivät voi olla syy- ja seuraussuhteessa. Esimerkissä 257 on syiden ja seurausten ketju, josta on aineistossa muitakin esimerkkejä.

**Koetinehdokas *результат*.** Manuaalisessa koetinhaussa produktiivisimmaksi syyn ja seurauksen koettimeksi osoittautui ilmaus *в результате* (tuloksena), joka koostuu prepositiosta *в* ja substantiivin *результат* prepositionaalimuodosta ja jota käytetään preposition tavoin ilmaisemaan syytä. Konkordanssihaku aineistosta *Bezop* hakuehdolla *в результате* tuotti 42 osumaa, joissa on mainittu ainakin yksi syy ja ainakin yksi seuraus. 26 osumassa sekä syy että seuraus kuuluu merenkulun erikoisalalle. Muissa osumissa syy- ja seurauskäsitteet ovat muiden erikoisalojen tai yleiskielen käsitteitä. 16 tapauksessa sekä syy että seuraus ovat tapahtumia:

- (259) - - в районе мыса доброй Надежды потерпели аварии несколько супертанкеров в результате разламывания корпуса. (Bezop: 100)

Esimerkissä 259 seurauskäsitteen nimitys *авария* (haveri) on ennen koetinehdokasta *в результате* ja syykäsitteen nimitys *разламывание корпуса* (rungon katkeaminen) välittömästi koetinehdokkaan jälkeen. Tapauksia, joissa toiminnan seurauksena on tapahtuma, on konkordanssilistassa viisi:

- (260) Теплоход ”Комсомолец Калмыкин” под прессом эксплуатационных служб вышел в штормовую погоду из порта с незакрепленным грузом длинномерного металла на твиндеках, погруженного без применения сепарации, в результате чего на качке произошло смещение груза и опрокидывание судна. (Bezop: 17)

Esimerkissä 260 toiminnan tai paremminkin toiminnan puuttumisen seurauksena on tapahtuma *смещение груза* (lastin siirtyminen), jonka seurauksena on toinen tapahtuma *опрокидывание судна* (aluksen kaatuminen). Tapahtumiin johtava syy voi lastin huonon sijoittelun ohella tosin olla lastin kiinnittämisen laiminlyönti tai päätös lähteä merelle huonosta säästä välittämättä, jolloin syitä lastin siirtymiselle voi olla monta. Joka tapauksessa kaikki mahdolliset syyt ovat ihmisen toimintaa. Toiminta on syynä tapahtumaketjuun, tapahtumaketju on syynä tapahtumaan tai olosuhde on syynä tapahtumaan kaikkiaan seitsemässä konkordanssilistan osumassa. Syy- tai seurauketjujen lisäksi syyt tai seuraukset voivat olla vaihtoehtoisia:

- (261) Известно, что подавляющее большинство пробоин происходит в результате столкновений или посадки на мель. (Bezop: 99)

Esimerkissä 261 reiän syy on joko yhteentörmäys tai karilleajo. Konkordanssilistassa on kuusi osumaa, joissa seuraukselle on vaihtoehtoisia syitä tai syyllä on vaihtoehtoisia seurauksia. Yhdessä tapauksessa seuraus on mahdollinen, mikä on ilmaistu modaalipuverbilla *мочь* (voida).

**Koetinehdokas *причина*.** Mahdollisista syyn ja seurauksen välisen funktiosuhteen substantiivikoettimista myös *причина* (syy) osoittautui produktiiviseksi tutkimusaineistossani. Konkordanssihakua hakuehdolla *причин*\* aineistosta Bezop tuotti 49 osumaa, joista valideja syyn ja seurauksen välisen funktiosuhteen konteksteja on 11. Lisäksi konkordanssilistassa on 18 kontekstia, joissa on mainittu seuraus, mutta ei syytä. Näistä konteksteista 12:ssa seuraus on jonkinlainen haveri. Muissakin osumissa on mainittu syitä ja seurauksia, mutta ne eivät kuulu merenkulun erikoisalalle.

Niissä valideissa osumissa, joissa on sekä syy että seuraus, seuraus on tavallisesti mainittu ennen koetinehdokasta ja syy tai syyt sen jälkeen. Syy- tai seurauskäsitteen nimityksen ja koetinehdokkaan välissä on kuitenkin tavallisesti useita käsitteen nimitykseen kuulumattomia sanoja. Kahdessa validissa osumassa syy tai seuraus on mainittu tekstissä niin paljon ennen koetinehdokasta, ettei se ole tulostunut konkordanssilistaan. Syitä tai seurauksia voi olla useita, kuten esimerkiksi 262, jossa aluksen menetyksen syitä ovat räjähdys ja tulipalo:

(262) Большое количество судов гибнет по причинам взрывов и пожаров. (Bezop: 23)

Seuraus on kaikissa valideissa esimerkeissä tapahtuma, mutta syy voi joskus olla myös muu kuin tapahtuma.

**Koetinehdokas *следствие*.** Manuaalisen koetinhaun perusteella kolmaskin substantiivi, *следствие* (seuraus), on todennäköinen syy- ja seuraussuhteen koetin. Konkordanssihakua hakuehdolla *\*следстви\** tuotti 26 osumaa, joista 10 on kelvollisia tapauksia koetinehdokkaan toiminnasta syy- ja seuraussuhteen koettimena. Esimerkissä 263 rungon pitkittäisjännitykset ovat mahdollinen repeämien syy:

(263) Учитывая большую длину балкера и огромные количества размещенных на нем грузов, при нарушении установленной технологии грузовых операций могут возникнуть запредельные напряжения в корпусе и как следствие того трещины. (Bezop: 110)

(264) Пожары в трюмах характерны медленным развитием, но сильным задымлением вследствие ограниченного газообмена. (Bezop: 83)

Muissa osumissa kerrotaan, että tapahtumalla tai toiminnalla on (vakavia) seurauksia, mutta ei kerrota, mitä nämä seuraukset ovat.

**Koetinehdokas *цель*.** Manuaalisen koetinhaun perusteella substantiivi *цель* (tarkoitus, tavoite) on mahdollinen venäläinen tarkoituksen koetin. Koettimena toimiessaan se on joko muodossa *целью* tai *в целях*. Konkordanssihakua hakuehdolla *целью/в целях* tuotti 17 osumaa, joista kuudessa koetinehdokas toimii tarkoituksen koettimena. Kaikissa valideissa osumissa tarkoituspierre liittyy toimintaan, jonka tarkoitus on toisen toiminnan toteuttaminen:

(265) Смыслом и главной целью борьбы за живучесть судна является спасение человеческих жизней. (Bezop: 64)



Esimerkissä 265 koetinehdokas liittyy aluksen pelastamiseen (борьба за живучесть судна), jonka ensisijainen tarkoitus on ihmishengen pelastaminen (спасение человеческих жизней). Muissa osumissa toiminta ja sen tarkoitus eivät liity merenkulun turvallisuuteen.

**Koetinehdokas *следовательно*.** Tapahtuman syytä ja seurausta sekä toiminnan tarkoitusta ilmaisevien mahdollisten verbi- ja substantiivikoettimien lisäksi manuaalinen haku tuotti adverbikoetinehdokkaan *следовательно* (niin muodoin, siis). Konkordanssihaku hakuehdolla *следовательно* aineistosta *Bezop* tuotti 26 osumaa, joista yhdeksän on validia tapausta koetinehdokkaan toiminnasta syy- ja seurauskäsitteiden välisen funktiosuhteen koettimena. Kaikissa valideissa tapauksissa syykäsitteen nimitys on pilkun ja konjunktion *и* tai *а* päässä välittömästi ennen koetinehdokasta ja seurauskäsitteen nimitys pilkun ja mahdollisesti preposition tai konjunktion päässä välittömästi koetinehdokkaan jälkeen.

(266) - - могло принести только к усилению эксплуатации его членов, к усталости и, следовательно, к снижению уровня безопасности, - - (Bezop: 16)

Esimerkissä 266 syykäsitteen nimitys *усталость* (väsymys) on ennen koetinehdokasta. Välissä on konjunktio *и* ja pilkku. Seurauskäsitteen nimitys *снижение уровня безопасности* (turvallisuuden heikkeneminen) on koetinehdokkaan jälkeen, välissä on vain prepositio *к*.

**Koetinehdokas *из-за*.** Konkordanssihaku hakuehdolla *из-за* (takia, vuoksi) aineistosta *Bezop* tuotti 13 osumaa, joista valideja syyn ja seurauksen välisen funktiosuhteen tapauksia on kahdeksan. Syykäsitteen nimitys on yhtä tapausta lukuun ottamatta välittömästi koetinehdokkaan jälkeen. Ainoassa poikkeustapauksessakin seurauskäsitteen nimitys on välittömästi ennen koetinehdokasta. Seuraus on aina tapahtuma, mutta syy voi olla ominaisuus, olotila tai olio, vaikka useimmiten syykin on tapahtuma, kuten esimerkiksi 267, jossa lastin siirtyminen (подвижка груза) on syynä toiseen tapahtumaan, vakavaan havariin (серьезная авария судна):

(267) Исследуя аварийную статистику отечественного флота за последние 15 лет, можно увидеть, что значительное число случаев серьезных аварий судов произошло из-за подвижки груза, - - (Bezop: 17)

**Yhteenveto.** Vahvimpia venäläisiä kausaalisen käsitesuhteen koettimia ovat konkordanssihakujen tulosten perusteella verbi *привести*, sen imperfektiivinen aspektipari *приводить* ja verbi *вызвать*, jotka olivat produktiivisia ja tuottivat vähän roskaa, sekä ilmaus *в результате* ja prepositio *из-за*. Prepositio *из-за* ei ollut erityisen produktiivinen, mutta osoittautui hyväksi syyn ja seurauksen koettimeksi, sillä yli puolet sen tuottamista osumista on valideja.

Venäläisistä syyn ja seurauksen substantiivikoettimista vahvimமாகsi osoittautui *в результате*. Konkordanssihaku aineistosta *Bezop* hakuehdolla *в результате* tuotti runsaasti osumia, joissa on mainittu ainakin yksi syy ja ainakin yksi seuraus. Osumista liki 2/3 on valideja. Myös substantiivit *причина*, *вследствие* ja *цель* sekä adverbi *следовательно* tuottivat paljon osumia tutkimukseni aineistosta, mutta osumista alle puolet on valideja.

Verbikoettiin liittyy tavallisesti modaaliapuverbi *мочь*. Kaikille syyn, seurauksen ja tarkoituksen koettimille on tyypillistä, että syitä ja seurauksia voi olla monta tai ne voivat olla vaihtoehtoisia.

### 7.5.7.2 Vaikuttavan tekijän venäläiset koettimet

Noudatan venäläisten koettimien tarkastelussa samaa jaottelua kuin suomalaisten koettimien tarkastelussa ja erotan vaikuttavan tekijän koettimet omaksi ryhmäksi (ks. jakso 7.4.7). Aiemmissa venäläisellä aineistolla tehdyissä tutkimuksissa ei ole mainittu vaikuttavan tekijän venäläisiä koettimia. Suomalaisella aineistolla tehty koetinhaku tuotti kuitenkin vaikuttavan tekijän verbi- ja substantiivikoettimia, joten oletin niitä löytyvän myös venäläisestä aineistosta. Venäläisellä artikkelilla *Eskos* tehty koe, jossa tuotettiin tunnettujen termien konkordanssit, ei tuottanut yhtään vaikuttavan tekijän koetinta, mutta manuaalinen haku aineistosta *Bezop* tuotti 42 osumaa, joissa on vaikuttava tekijä. Manuaalisen haun perusteella vaikuttavan tekijän venäläisiä koettimia ovat verbit *сказываться* (vaikuttaa), *воздействовать* (vaikuttaa), *зависеть* (riippua) ja *влиять* (vaikuttaa), substantiivit *влияние* (vaikutus), *воздействие* (vaikutus) ja *фактор* (tekijä) sekä ilmaus *за счет* (syystä). Näistä koetinehdokkaista muodostetuilla hakuehdoilla tein konkordanssihaut, joiden tulokset ovat taulukossa 55.

TAULUKKO 55. Vaikuttavan tekijän terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
сказываться	<i>сказыва*</i>	4	3	75	hyvä
воздействовать	<i>воздействиу*</i>	3	3	100	hyvä
зависеть от	<i>зависит от</i>	17	8	47	hyvä
влиять	<i>влияет/влияют</i>	5	4	80	hyvä
влияние	<i>влияни*</i>	31	20	65	vahva
воздействие	<i>воздействи*</i>	14	5	36	hyvä
зависимость	<i>в зависимости от</i>	15	6	40	hyvä
фактор	<i>фактор</i>	30	6	20	heikko
за счет	<i>за счет</i>	14	11	79	vahva

<sup>a</sup> Ks. jakso 7.3.10.

**Koetinehdokas *сказываться*.** Tein ensin konkordanssihaun verbillä *сказываться* (vaikuttaa). Hakuehto *сказыва\** tuotti neljä osumaa, joista kolme on validia tapausta koetinehdokkaan toiminnasta vaikuttavan tekijän koettimena:

(268) Развившаяся система подставных флагов сказывается на уровне безопасности мореплавания крайне отрицательно, - - (Bezop: 15)

Esimerkissä 268 vaikuttava tekijä on система подставных флагов (mukavuuslippujärjestelmä), jonka vaikutuskohde on безопасность мореплавания (merenkulun turvallisuus). Esimerkissä on ilmaistu myös vaikutuksen suunta, joka on negatiivinen. Verbi *сказываться* ei ole produktiivinen, mutta ei myöskään tuota paljoa roskaa. Verbin käyttökelpoisuutta koettimena olisi kuitenkin testattava suuremmalla aineistolla.

**Koetinehdokas *воздействовать*.** Seuraavaksi tein konkordanssihaun hakuehdolla *воздействию\**, joka kattaa verbin *воздействовать* (vaikuttaa) preesensmuodot. Haku tuotti kolme osumaa, joista jokainen on validi tapaus koetinehdokkaan toiminnasta vaikuttavan tekijän koettimena:

- (269) В условиях морской перевозки, когда на грузы воздействуют качка, динамические удары волн, вибрация судна, статическое давление штабеля высотой до 10 м, закрытая атмосфера трюма, свойства этих грузов могут проявиться самым неожиданным образом и явиться причиной возникновения аварий судов и гибели членов экипажа. (Bezop: 17)

Esimerkissä 269 vaikutuksen kohde on lasti (груз), mutta koetinverbi *воздействовать* on monikossa, koska vaikuttavia tekijöitä on monta: keinunta (качка), aallon iskukuorma (динамические удары волн), värinä (вибрация судна), lastin aiheuttama paino (статическое давление штабеля) ja suljettu ilmankierto (закрытая атмосфера трюма). Tässä esimerkissä ei suoraan sanota, onko vaikutus positiivinen vai negatiivinen.

**Koetinehdokas *зависеть от*.** Manuaalisen haun perusteella verbi *зависеть* (riippua) on mahdollinen vaikuttavan tekijän venäläinen koetin. Se näyttäisi toimivan koettimena yhdessä preposition *от* kanssa. Konkordanssihaku hakuehdolla *зависит от* aineistosta Bezop tuotti 17 osumaa, joista kahdeksan on validia tapausta koetinehdokkaan toiminnasta vaikuttavan tekijän koettimena. Esimerkissä 270 vaikuttava tekijä on ominaislasti-tilavuus (удельный погрузочный объем), jonka vaikutuksen kohde on ruuman täyttöaste (степень загрузки трюма):

- (270) Коэффициент проницаемости трюмов на балкерах нового поколения будет зависеть от степени их загрузки, что в свою очередь зависит от удельного погрузочного объема (УПО) перевозимого груза. (Bezop: 110)

Vaikuttavaa tekijää nimeävä termi on kaikissa valideissa tapauksissa välittömästi koetinehdokkaan *зависит от* jälkeen. Koetinehdokas onkin käyttökelpoisin, jos siihen voi hakuehdossa yhdistää tunnetun, vaikuttavaa tekijää nimeävän termin, jolloin kelpaamatomat osumat karsiutuvat osumalistasta. Koetinehdokkaan käyttökelpoisuutta heikentää se, että käsitesuhteen toinen osapuoli, vaikutuksen kohde, on viidessä tapauksessa kahdeksasta ilmaistu pronomiinilla ja varsinaista kohteen nimitystä on etsittävä edeltävästä tekstistä.

**Koetinehdokas *влиять*.** Manuaalisen haun perusteella verbi *влиять* (vaikuttaa) on koettimena toimiessaan yksikön 3. persoonan preesensmuodossa *влияет* tai monikon 3. persoonan preesensmuodossa *влияют*. Hakuehdolla *влияет/влияют* tehty konkordanssihaku tuotti aineistosta Bezop viisi osumaa, joista neljä on valideja vaikuttavan tekijän tapauksia. En ole hyväksynyt osumaa kelpvolliseksi, jos vaikutuksen kohde puuttuu korpusohjelman tuottamasta kontekstista. Verbimuoto *влияет* on koettimena kahdessa ja muoto *влияют* samoin kahdessa validissa osumassa:

- (271) Если рассматривать взаимодействие судна, груза и экипажа с позиций эргономики и инженерной психологии, то судно, являясь элементом системы "человек-судно-среда", в свою очередь выступает как сложная инженерная система, включающая в себя целый ряд судовых систем, надежность и безопасность которых существенно влияют на безопасность плавания судна вообще. (Bezop: 41)

Esimerkissä 271 vaikutuksen kohteen nimitys *безопасность плавания судна* (aluksen käyttöturvallisuus) on välittömästi koettimen jälkeen, välissä on vain prepositio *на*. Vaikuttavia tekijöitä ovat alusjärjestelmien luotettavuus (надежность судовых систем) ja alusjärjestelmien turvallisuus (безопасность судовых систем), jotka kuitenkin on ilmaistu tekstissä elliptisesti. Tässä esimerkissä verbiin liittyy prepositio *на*, mutta konkordanssiliistassa on yksi osuma, jossa verbi toimii koettimena ilman prepositiota. Osumia on kuitenkin niin vähän, ettei niiden perusteella voi päätellä, kuinka usein prepositio liittyy verbiin niissä tapauksissa, joissa verbi toimii koettimena.

**Koetinehdokas *влияние*.** Substantiivi *влияние* (vaikutus) osoittautui produktiivisemmaksi kuin verbi *влиять*, sillä hakuehdon *влияни\** tuottamista 31 osumasta 20 on valideja osumia substantiivin toiminnasta vaikuttavan tekijän koettimena:

- (272) Груз как элемент транспортной системы оказывает большое влияние на безопасность плавания судна. (Bezop: 17)

Esimerkissä 272 vaikuttavan tekijän nimitys *груз* (lasti) on virkkeen alussa kuusi sanaa ennen koetinta, mutta vaikutuksen kohteen nimitys *безопасность плавания судна* (aluksen käyttöturvallisuus) on välittömästi koettimen jälkeen. Tässä esimerkissä substantiiviin liittyy välittömästi prepositio *на*, mutta prepositio saattaa myös puuttua, kuten viidessä validissa tapauksessa, tai substantiivin ja preposition välissä voi olla 2–4 sanaa, kuten kahdeksassa tapauksessa. Hakuehdon rajaaminen preposition avulla roskan vähentämiseksi ei siis kannata, koska näin menetettäisiin suurin osa valideista osumista, vaikka samalla saataisiinkin karsittua kolme neljäsosaa roskaosumista.

**Koetinehdokas *воздействие*.** Verbin *воздействовать* lisäksi myös samaa kantaa oleva substantiivi *воздействие* (vaikutus) on manuaalisen haun perusteella mahdollinen koetin. Hakuehto *воздействи\** tuotti 14 osumaa, joista viisi on validia:

- (273) К числу опасностей, сопутствующих перевозке сжиженных газов, относят: взрыва- и пожароопасность; токсичность, заключающуюся во вредном воздействии паров и жидкостей на дыхательные пути, - - (Bezop: 104)

Esimerkissä 273 on kaksi vaikuttavaa tekijää: höyry (пар) ja neste (жидкость), jotka vaikuttavat negatiivisesti hengitysteihin (дыхательные пути).

**Koetinehdokas *зависимость*.** Aiemmin totesin, että verbin *воздействовать* lisäksi substantiivi *воздействие* toimii koettimena. Niinpä tein konkordanssihaun verbin *зависеть* lisäksi substantiivilla *зависимость* (riippuvuus). Hakuehto *в зависимости от* (riippuen) tuotti 15 osumaa, joista kuusi on validia tapausta koetinehdokkaan toiminnasta vaikuttavan tekijän koettimena. Esimerkissä 274 virkkeen alussa on vaikutuksen kohteen nimitys *тушение пожаров в трюмах сухогрузных судов* (palonsammutus kuivarahtialuksen ruumassa) ja koettimen jälkeen vaikuttavien tekijöiden nimitykset *под груза* (tavaralaji) ja *свойства груза* (lastin ominaisuudet):

- (274) Тушение пожаров в трюмах сухогрузных судов производят в зависимости от рода груза и его свойств. (Bezop: 83)

Kaikissa hakuehdon *в зависимости от* tuottamissa valideissa osumissa vaikuttava tekijä on ilmaistu välittömästi koettimen jälkeen, mutta vaikutuksen kohteen nimitys on aina paljon ennen koetinta ja usein vaikeasti paikallistettavissa.

**Koetinehdokas *фактор*.** Käsinoiminnan perusteella myös substantiivi *фактор* (tekijä) on mahdollinen vaikuttavan tekijän koetin. Haku aineistosta *Bezop* hakuehdolla *фактор*\* tuotti 30 osumaa, joista kuusi on validia esimerkiksi koetinehdokkaan toiminnasta vaikuttavan tekijän koettimena:

- (275) Управляемая система может выйти из заданного состояния под воздействием трех основных факторов – внутренних отказов, возмущений внешней среды и ошибок человека. (Bezop: 27)

Esimerkissä 275 on kolme vaikuttavaa tekijää: järjestelmävika (внутренний отказ), ulkoinen häiriö (возмущение внешней среды) ja ihmisen virhetoiminta (ошибка человека). Tekijöiden vaikutuksen kohde on järjestelmä (система). Koetinehdokkaan tuottamat tapaukset eivät kuitenkaan ole kovin selkeitä, ja ne voi tulkita muullakin tavalla, kuin olen ne tulkinnut. Esimerkki 275 on ehkä selkeimmin tulkittavissa esimerkiksi vaikuttavasta tekijästä, koska virkkeessä on myös substantiivi *воздействие*, joka osoittaa, että luetellut tekijät ovat vaikuttavia tekijöitä.

**Koetinehdokas *за счет*.** Vaikuttavan tekijän mahdollinen koetin, ilmaus *за счет* (syystä), tuotti konkordanssihaulla 14 osumaa. Niistä 11 on kelvollisia tapauksia funktiosuhteesta, jossa on vaikuttava tekijä:

- (276) При получении пробоины в грузовых танках, если танкер в грузу, происходит не только затопление танка водой (полное или частичное за счет замещения груза), но и вытекание груза за борт, что является экологическим бедствием. (Bezop: 99)

Esimerkissä 276 vaikuttavaa tekijää nimeävä termi *замещение груза* (lastin siirtyminen) on välittömästi koetinehdokkaan jälkeen, kuten muissakin valideissa tapauksissa. Vaikuttavan tekijän kohteena olevan käsitteen nimitys on yleensä välittömästi ennen koetinehdokasta. Esimerkissä 276 välittömästi ennen koetinehdokasta on kohteena olevan käsitteen elliptinen nimitys *частичное*.

**Yhteenveto.** Vahvimpia venäläisiä vaikuttavan tekijän koettimia ovat konkordanssihakujen tulosten perusteella substantiivi *влияние* ja ilmaus *за счет*, jotka ovat produktiivisia eivätkä tuota ylen määrin roskaa. Verbikoettimista produktiivisin on *зависеть от*, jonka tuottamista osumista noin puolet on valideja. Vaikuttavaa tekijää nimeävä termi on aina välittömästi verbin jälkeen, joten parhaaseen tulokseen päästään, jos hakuehtoon voi liittää termin. Verbit *воздействовать*, *сказываться* ja *влиять* eivät ole yhtä produktiivisia, mutta eivät myöskään tuota paljon roskaa. Substantiivit *воздействие* ja *влияние* osoittautuivat produktiivisemmiksi kuin vastaavat verbit *воздействовать* ja *влиять*. Koettimien listaan voi lisätä myös substantiivin *зависимость*, joka koettimena toimiessaan on muodossa *в зависимости от*. Vaikuttavaa tekijää nimeävä termi on aina välittömästi koettimen jälkeen, joten tälläkin koettimella haku tarkentuu, jos sen voi tehdä yhdessä tunnetun termin kanssa. Sen sijaan substantiivi *фактор* osoittautui heikoksi vaikuttavan tekijän koettimeksi, sillä sen ilmaisemaan suhteeseen liittyy tulkinnanvaraisuutta. Tyypillistä on, että vaikuttavaa tekijää nimeävä termi on tekstissä välittömästi koettimen jälkeen, koska moni niistä käyttäytyy preposition tavoin. Joihinkin koettimiin liittyy monta vaikuttavaa tekijää, kuten koettimeen *воздействовать*.

### 7.5.7.3 Toimintasuhteiden venäläiset koettimet

Suomalaisten koettimien käsittelyssä jaottelin toimintasuhteet Nuoppota (2005) mukaillen menetelmä-, väline-, keino-, agenttiivi- ja objektiisuhteeseen. Noudatan venäläisten koettimien tarkastelussa samaa jaottelua. Keino voi liittyä myös tapahtumakäsitteeseen, silloin kun keinoa käytetään toivotun tapahtuman edistämiseen tai ei-toivotun tapahtuman estämiseen. Käsittelen näitä käsitesuhteita samassa jaksossa, koska kaikissa niissä suhteen toinen käsite on toiminta- tai tapahtumakäsite, jonka toteutumista keino, menetelmä tai väline edistää tai estää. Samassa validissa osumassa voi olla esimerkiksi keinon ja tapahtuman sekä välineen ja tapahtuman välinen suhde.

Manuaalinen koetinhaku tuotti yhden mahdollisen menetelmän koettimen, substantiivin *метод* (menetelmä). Toimintakäsitteen ja sen toteuttamiseen käytettävän välineen koetinehdokkaista ovat haun perusteella verbit *применять* (soveltaa, käyttää), *использовать* (käyttää), *служить* (palvella) ja *предназначить* (määrätä, tarkoittaa) sekä substantiivi *помощь* (apu). Keinosuhteen koetinehdokkaista ovat verbi *достигать* (saavuttaa) ja substantiivit *средство* (keino, väline), *способ* (keino) ja *путь* (tie). Toimijan mahdollisia koettimia ovat verbit *осуществлять* (toteuttaa) ja *обеспечивать* (taata, turvata). Koetinehdokkaista muodostetuilla hakuehdoilla tein konkordanssihaun, joiden tulokset olen koonnut taulukkoon 56.

TAULUKKO 56. Toimintasuhteen terminologiset koettimet venäjänkielisessä aineistossa

Koetinehdokas	Hakuehto	Osumat f	Validit f	osumat %	Arviointi <sup>a</sup>
метод	<i>методом/методами</i>	11	8	73	vahva
применять	<i>применяют</i> <i>применяется/применяются</i>	17 25	12 9	71 36	vahva hyvä
использовать	<i>используют</i> <i>используется/используются</i>	11 8	8 2	73 25	vahva heikko
служить	<i>служит/служат/служащ*</i>	13	9	69	vahva
предназначить	<i>предназначен*</i>	10	7	70	vahva
средство	<i>средств*</i>	72	11	15	heikko
достигать	<i>достига*</i>	9	4	44	heikko
способ	<i>способ/способом</i>	18	7	39	hyvä
путь	<i>путем</i>	16	11	67	vahva
помощь	<i>с помощью</i> <i>при помощи</i>	8 3	6 3	75 100	hyvä hyvä
осуществлять	<i>осуществля*</i>	17	5	29	heikko
обеспечивать	<i>обеспечива*</i>	20	4	20	heikko

<sup>a</sup> Ks. jakso 7.3.10.

**Menetelmäsuhde.** Substantiivi *метод* (menetelmä) on manuaalisen haun perusteella mahdollinen menetelmäsuhteen koetin. Konkordanssihaku aineistosta *Bezop* hakuehdolla *метод\** tuotti 32 osumaa, joista kahdeksan on validia tapausta koetinehdokkaan toiminnasta toimintakäsitteen ja menetelmän välisen funktiosuhteen ilmaisimena. Esimerkissä 277 on mainittu kaksi menetelmää, joita käytetään vakavuuslaskelmien tekoon.



Menetelmät ovat täyden uppouman menetelmä (метод приема груза) ja vakiouppoumamenetelmä (метод постоянного водоизмещения):

- (277)      Ранее указывалось, что все расчеты по непотопляемости производятся либо методом приема груза (отсеки первой и второй категории), либо методом постоянного водоизмещения (отсеки третьей категории). (Bezop: 91)
- (278)      Снижения уровня аварийных расходов можно добиться различными методами: внедрением новых технологий, техники, конструктивных решений, организационных мероприятий и др. (Bezop: 14–15)

Valideissa osumissa substantiivi on joko yksikön tai monikon instrumentaalissa. Tarkentamalla hakuehtoa muotoon *методом/методами* saadaan osumalistaan 11 osumaa, joista kahdeksan eli 73 prosenttia on kelvollisia. Toinen mahdollisuus on tehdä haku koetinehdokkaalla yhdistyneenä toimintakäsitteen nimitykseen, joka on melkein kaikissa valideissa osumissa välittömästi koetinehdokkaan jälkeen.

**Вälinesuhde.** Manuaalisen haun perusteella verbit *применять* (soveltaa, käyttää), *использовать* (käyttää), *служить* (palvella) ja *предназначить* (määrätä, tarkoittaa) sekä ilmaukset *с помощью* ja *при помощи* (avulla) ovat todennäköisiä toimintakäsitteen ja sen toteuttamiseen käytettävän välineen koettimia. Näiden koetinehdokkaiden testaamiseksi tein konkordanssihaut aineistosta *Bezop* hakuehdoilla *применяют, применяется/применяются, используют\*, служит/служат/служащ\*, предназначен\*, с помощью* ja *при помощи*.

**Koetinehdokas *применять*.** Manuaalisen haun perusteella toiminnan ja toimintaan käytetyn välineen välisen suhteen mahdollinen verbikoetin *применять* (soveltaa, käyttää) on koettimena toimiessaan muodossa *применяют* tai passiivimuodossa *применяется* tai *применяются*. Konkordanssihaku aineistosta *Bezop* hakuehdolla *применяют* tuotti 17 osumaa, joista 12:ssa koetinehdokas osoittaa toiminnan ja sen toteuttamiseen käytettävän välineen tai joskus keinon välistä funktiosuhdetta. Valideissa tapauksissa koetinehdokkaaseen liittyy aina joko prepositio *для* tai ilmaus *в качестве*. Yhdessä tapauksessa koetinehdokkaaseen liittyy ilmaus *с целью*:

- (279)      Насыщенный водяной пар применяют для тушения пожаров газообразных, жидких и твердых веществ на небольших площадях в закрытых помещениях объемом не более 500 м<sup>3</sup>. (Bezop: 80)
- (280)      Воздушно-механическую пену средней и высокой кратности применяют в качестве основного средства тушения пожаров нефтепродуктов. (Bezop: 81)
- (281)      Инертные газы (азот, аргон, гелий, дымовые или отработанные газы и др.) применяют в основном с целью понижения концентрации кислорода в очаге горения. (Bezop: 81)

Esimerkissä 279 toiminta on sammutus (тушение) ja siihen käytettävä väline vesihöyry (водяной пар). Validit tapaukset voi tosin tulkita toisinkin: koetinehdokas osoittaa yhtä hyvin olion käyttötarkoitusta. Tämä tulkinta on jopa todennäköisempi, jos oliokäsitteen nimitys on tekstissä ennen koetinta ja toimintakäsitteen nimitys koettimen jälkeen, kuten esimerkeissä 279–281.

Olio- ja toimintakäsitteiden nimitykset ovat välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen kahdeksassa tapauksessa 12:sta. Kolmessa tapauksessa välissä

on jompikumpi ilmauksista *в основном* tai *с успехом*, ja yhdessä tapauksessa käsitteen nimitys on korvattu pronomiinilla. Neljässä ei-validissa osumassa on kerrottu olion käyttöpaikka, mutta ei toimintaa, jonka toteuttamisen väline tai keino olio voisi olla.

Hakuehdolla *применяется/применяются* konkordanssihaku aineistosta *Bezop* tuotti 25 osumaa, joissa kaikissa koetinehdokas liittyy olion käyttötarkoitukseen, käyttöpaikkaan tai käyttöaikaan. Yhdeksässä tapauksessa koetinehdokas on toiminnan ja toiminnan välineen koettimena:

(282) Пена широко применяется для тушения горящих нефтепродуктов - - (Bezop: 59)

(283) Переносные электрические дымососы применяются для вентиляции защищаемых помещений. (Bezop: 62)

Esimerkissä 282 koetinehdokas *применяется* ilmaisee toiminnan тушение (sammutus) ja siihen käytettävän välineen пена (vaahto) välistä suhdetta. Toiminnan ja toiminnan välineen suhteen koettimena yksikkömuoto näyttäisi olevan käyttökelpoisempi, sillä sen tuottamista osumista yli puolet on valideja, kun taas monikkomuodon *применяются* tuottamista osumista vain joka neljäs on validi. Koetinehdokkaaseen liittyy kaikissa valideissa tapauksissa prepositio *для*. Kelpaamattomista osumista puuttuu toimintakäsitteen nimitys. Niissä puhutaankin tavallisesti olion käyttöpaikasta käyttötarkoituksen sijasta.

Hakuehdon *применяются* tuottamista neljästä validista osumasta kolmessa toimintakäsitteen nimitys on mainittu tekstissä niin paljon aikaisemmin, ettei se ole tulostunut konkordanssilistaan. Verbimuoto *применяют* näyttäisi toimivan toimintakäsitteeseen funktiosuhteen kautta liittyvän välineen koettimena paremmin kuin passiivimuodot *применяется* ja *применяются*.

**Koetinehdokas *использовать*.** Verbi *использовать* (käyttää) toimii koettimena hyvin samalla tavoin kuin verbi *применять*. Koettimena ollessaan se on monikon 3. persoonan muodossa *используют*, jolla konkordanssihaku tuotti 11 osumaa. Niistä kahdeksassa koetinehdokas osoittaa toiminnan ja siihen käytettävän välineen välistä funktiosuhdetta.

Passiivimuoto *использоваться* ei ollut yhtä produktiivinen kuin verbi *применять* tai verbi *использовать* aktiivissa käytettynä. Konkordanssihaku hakuehdolla *используется/используются* aineistosta *Bezop* tuotti kahdeksan osumaa, joista kuitenkin vain kahdessa on toimintakäsitteeseen liittyvän keinon tai välineen nimitys.

Verbi *использовать* toimii välineen koettimena hyvin samalla tavalla kuin verbi *применять*. Kummankin verbin konkordanssihaun tuloksissa on joko toiminta ja sen toteuttamiseen käytetty väline tai olion toimiminen jossakin roolissa. Yhteistä on myös se, että koetinverbiin liittyy prepositio *для* tai joissakin tapauksissa *при* (*на, как средство*), silloin kun olio on toiminnan toteuttamisen väline, kuten esimerkissä 284. Prepositio ei kuitenkaan aina liity välittömästi koetimeen, vaan välissä saattaa olla 1–3 sanaa, kuten esimerkissä 285. Prepositio voi olla myös ennen koetinta, kuten esimerkissä 286. Silloin kun olion tarkoitus on toimia roolissa, kuten esimerkissä 287, koetinehdokkaaseen liittyy ilmaus *в качестве*:

(284) Пену используют для создания слоя на поверхности горящего вещества. (Bezop: 81)

(285) Дымосигнальные установки применяются повсеместно и используются в основном для обнаружения пожара в грузовых помещениях и кладовых. (Bezop: 56)

(286) Тушение пожаров в трюмах сухогрузных судов производят в зависимости от рода груза и его свойств. При возгорании опасных грузов используют огнегасительные средства, регламентированные Правилами МОПОГ. (Bezop: 83)

(287) В качестве плавкой вставки используется сплав Вуда, - - (Bezop: 56)

Käsitteiden nimitysten ja koetinehdokkaan keskinäinen järjestys tekstissä vaihtelee. Seitsemässä tapauksessa oliokäsitteen nimitys on välittömästi ennen koetinehdokasta tai välittömästi sen jälkeen ja kolmessa tapauksessa toimintakäsitteen nimitys on välittömästi ennen koetinehdokasta.

**Koetinehdokas *служить*.** Verbi *служить* (palvella) on kolmas venäläinen verbi-koetinehdokas, joka manuaalisessa koetinhaussa osoittautui todennäköiseksi toiminnan ja välineen välisen suhteen koettimeksi. Manuaalisen haun perusteella muotoiltu hakuehto *служит/служат/служащ\** tuotti 13 osumaa, joista yhdeksässä koetinehdokas toimii välinesuhteen koettimena:

(288) Нижние коуши и кренгельс служат для крепления подкильных концов. (Bezop: 52)

(289) Учебный пластырь (пластырь Макарова) имеется на всех судах и служит для тренировки экипажа при отработке действий по борьбе с водой. (Bezop: 53)

Esimerkissä 289 opetusvuotomatto (учебный пластырь) toimii miehistön harjoittelun (тренировка экипажа) välineenä. Silloin kun olion tarkoitus on toiminnan toteuttaminen, koetinehdokasverbiin liittyy kuudessa tapauksessa prepositio *для*, kuten esimerkeissä 288 ja 289, ja yhdessä tapauksessa ilmaus *как средство*. Prepositio ja sitä seuraava toimintakäsitteen nimitys on tavallisesti välittömästi verbin jälkeen, ja välinettä nimeävä termi tai sen sijasta käytetty pronomini on tavallisesti välittömästi ennen koetinehdokasta. Verbin yksikkö- ja monikkomuoto tuottavat suunnilleen yhtä monta osumaa ja validia tapausta. Verbi *служить* ei ole produktiivisin toimintasuhteen koetiverbeistä, mutta sen tarkkuus on varsin hyvä, sillä noin kaksi kolmasosaa osumista on valideja.

**Koetinehdokas *предназначить*.** Manuaalinen koetinhaku tuotti yhden mahdollisen verbikoettimen, joka välinesuhteen koettimena toimiessaan on aina passiivin partisiipin preteritimuodossa. Tämä on verbi *предназначить* (määrätä, tarkoittaa), jonka partisiippimuoto *предназначенный* toimii manuaalisen haun perusteella koettimena. Konkordanssihakuehdolla *предназначен\** aineistosta Bezop tuotti 10 osumaa, joista seitsemässä koetinehdokas osoittaa toiminnan ja toiminnan tekemiseen käytettävän oliokäsitteen suhdetta. Oliokäsitettä nimeävä termi on aina ennen koetinehdokasta. Neljässä tapauksessa se on välittömästi ennen sitä, kuten esimerkissä 291, ja kolmessa tapauksessa välissä on 1–2 sanaa, kuten esimerkissä 290:

(290) Танкер – это судно, предназначенное для перевозки жидких грузов, имеющих свободную поверхность. (Bezop: 98)

(291) - - система порошкового тушения предназначена для борьбы с пожаром в районе грузовой зоны; - - (Bezop: 105)

Esimerkissä 290 koetinehdokas ilmaisee toiminnan, nestemäisten irtolastien kuljetuksen (перевозка жидких грузов, имеющих свободную поверхность), ja siihen käytettävän

välineen, säiliöaluksen (танкер), välistä suhdetta. Kahdessa tapauksessa termin tilalla välittömästi ennen koetinedokasta on pronomini ja itse termi on mainittu aiemmin tekstissä. Koetinedokkaaseen liittyy yhtä poikkeusta lukuun ottamatta kaikissa osumissa prepositio *для*, joka on yleensä välittömästi koetinedokkaan jälkeen. Neljässä tapauksessa tarkoituspieriteisiin liittyy rajoittava lisäpiirre. Kahdessa tapauksessa se on olosuhde ja kahdessa paikka, kuten esimerkissä 291. Siinä koetinedokas *предназначена* ilmaisee palontorjunnan (борьба с пожаром) ja jauhesammutusjärjestelmän (система порошкового тушения) välistä suhdetta, johon liittyy rajoittava lokaalinen piirre lastitila (в районе грузовой зоны).

**Koetinedokas *помощь*.** Manuaalinen koetinhaku osoitti, että ilmaukset *с помощью* (avulla) ja *при помощи* (avulla) ovat todennäköisiä toiminnan ja sen toteuttamiseen käytettävän keinon tai välineen välisen funktiosuhteen koettimia. Konkordanssihaku hakuehdolla *с помощью* tuotti aineistosta *Bezop* kahdeksan osumaa, joista kuusi on validia tapausta koetinedokkaan toiminnasta välineen koettimena, ja hakuehdolla *при помощи* haku tuotti kolme osumaa, joista jokainen on kelvollinen. Esimerkissä 292 koetinedokas toimii koettimena välineiden eli puutavaran (аварийный брус) ja jatkettavan tuen (раздвижной упор) sekä toiminnan eli vahvistamisen (подкрепление) välillä:

(292) Подкрепление производится с помощью аварийных брусьев и раздвижных упоров.  
(Bezop: 77)

(293) По мере накопления вода удаляется за борт при помощи осушительной системы.  
(Bezop: 49)

Esimerkissä 293 väline on kuivatusjärjestelmä (осушительная система), jonka avulla tapahtuma eli veden poistuminen yli laidan (вода удаляется за борт) toteutuu. Kaikissa yhdeksässä validissa tapauksessa välinettä nimeävä termi on tekstissä välittömästi koetinedokkaan *с помощью* tai *при помощи* jälkeen.

**Keinosuhde.** Artikkelista *Eskos* löytyi tunnettujen termien konkordanssihaulla yksi keinosuhteen koetinedokas: *являются средством* (olla keino). Sen lisäksi manuaalissa koetinhaussa mahdolliseksi toimintakäsitteen ja sen toteuttamiseen käytettävän keinon koettimiksi osoittautuivat verbi *достигать* (saavuttaa) ja substantiivit *средство* (keino, väline), *способ* (keino) ja *путь* (tie), erityisesti sen instrumentaalimuoto *путем* (avulla). Näiden koetinedokkaiden testaamiseksi tein konkordanssihaut aineistosta *Bezop* hakuehdoilla *достига\**, *средств\**, *способ\** ja *путем*.

**Koetinedokas *достигать*.** Toimintasuhteen mahdollisista verbikoettimista vähiten osumia tuotti verbi *достигать* (saavuttaa). Se tuotti aineistosta *Bezop* hakuehdolla *достига\** 9 osumaa, joista valideja keinosuhteen tapauksia on neljä. Niissä keino on selkeästi ilmaistu tekstissä välittömästi tai korkeintaan kaksi sanaa koetinedokkaan jälkeen, mutta toimintakäsitetä, johon keino on funktiosuhteessa, on vaikea hahmottaa edeltävästä virkkeestä tai lauseesta:

(294) Танкер, имея 80-90% площади свободной поверхности жидкости от площади действующей ватерлинии, неминуемо должен был бы опрокинуться, если бы не был конструктивно защищен от этой опасности. Это достигается установкой продольных водонепроницаемых переборок, --(Bezop: 98)

Esimerkissä 294 keino vesitiiviiden välilaipioiden asennus (установка продольных водонепроницаемых переборок) on ilmaistu välittömästi koetinedokkaan jälkeen, mutta kaatuminen (опрокинуться), joka keinon avulla pyritään estämään, on kaukana edeltävässä kontekstissa. Tutkimukseni aineistossa koetinedokas *достигать* ei ole vahva keinosuhteen koetin, koska se tuotti vähän osumia, joista alle puolet on valideja.

**Koetinedokas *средство*.** Tunnettujen termien konkordanssihaulla artikkelista *Eskos* löytynyttä keinosuhteen koetinedokasta *являться средством* (olla keino) käytetään artikkelissa kaksi kertaa ilmaisemassa keinoa toteuttaa merenkulun turvallisuuden ylläpito (обеспечение безопасности мореплавания). Esimerkissä 295 tämä keino on alusliikennepalvelu (СУДС):

- (295) СУДС общепризнанно являются важнейшим средством обеспечения безопасности мореплавания на акваториях портов и на подходах к ним, повышения эффективности транспортного процесса. (Eskos: 15)

Alusliikennepalvelun (СУДС) voidaan kuitenkin ajatella olevan yhtä hyvin toiminnan toteuttamisen väline kuin keino. Aineistosta *Bezop* konkordanssihaku hakuehdolla *средством* tuotti kaksi osumaa, joista toisessa substantiivi ilmaisee keinoa:

- (296) - - наиболее эффективным средством борьбы за живучесть является не только должная организация службы и уровень профессиональной подготовленности экипажа для действий в кризисной ситуации, но главным образом предотвращение ее. (Bezop: 49)

Esimerkissä 296 substantiivi *средство* ilmaisee keinoa, jolla pyritään pelastamaan alus. Koska hakuehto tuotti vain kaksi osumaa, se on luultavasti liian suppea. Konkordanssihaku laajemmalla hakuehdolla *средств\** tuotti 72 osumaa, joista 11 on kelvollista tapausta koetinedokkaan toiminnasta keinon tai monessa tapauksessa paremminkin välineen koettimena. Lisäksi yhdeksässä validissa tapauksessa keinoa nimeävä termi on konkordanssihaun tulostusalueen ulkopuolella tai keinoa ei ole mainittu. 10 tapauksessa toimintakäsitteen nimitys on tekstissä välittömästi koetinedokkaan jälkeen, kuten esimerkiksi 296 pelastaminen (борьба за живучесть). Kun koetinedokasta käytetään yhdessä toimintakäsitetä nimeävän termin kanssa, saadaan roskaa määrää pienennettyä. Näin menetetään vain yksi osuma, jossa koetinedokkaan ja toimintakäsitteen nimityksen välissä on nimitykseen kuulumattomia sanoja.

**Koetinedokas *способ*.** Keinon mahdollisista substantiivikoettimista eniten osumia tuotti *способ* (keino). Hakuehdolla *способ\** se tuotti 45 osumaa, joista 11 on kelvollisia tapauksia substantiivin toiminnasta keinon koettimena. 11 osumassa hakuehdon täyttää substantiivin *способ* sijasta substantiivi *способность*, ja kahdessa on adjektiivi *способный* pitkässä tai lyhyessä muodossa. Näistä osumista yksikään ei ole validi merenkulun toimintakäsitteen ja keinon välisen funktiosuhteen tapaus. Jos hakuehto tarkennetaan muotoon *способ/способом*, osumien määrä vähenee 18 osumaan. Substantiivin *способ* monikon genetiivimuoto *способов* on jätetty pois hakuehdosta, sillä se tuottaa vain yhden osuman, joka tosin on tässä aineistossa validi. Osumalistalla on 10 tapausta, joissa substantiivi *способ* on termin osa eikä koetin, kuten esimerkiksi 297, jossa substantiivi on osa termiä roro-lastausmenetelmä (горизонтальный способ погрузки):

- (297) На судах с горизонтальным способом погрузки деление на отсеки производится путем установки горизонтальных водонепроницаемых переборок, служащих грузовыми палубами и платформами. (Bezop: 49)

Esimerkissä 298 koetinedokkaan osoittama keino on sähkö- tai kaasuhitsauksen käyttö (применение электро- или газосварки). Sen avulla toteutettava toiminta on putkivian korjaus (исправление повреждений трубопроводов).

- (298) Самым надежным способом исправления повреждений трубопроводов является применение электро- или газосварки, - - (Bezop: 79)

Aiemmin totesin, että hakuehdon *средств\** tuottamaa roskaa voidaan huomattavasti vähentää hakemalla valideja konteksteja yhdessä toimintakäsitteen nimityksen kanssa, sillä valideissa osumissa tämä on tekstissä yleensä välittömästi koetinedokkaan jälkeen. Sama pätee substantiiviin *способ*, joka on viidessä validissa tapauksessa välittömästi ennen toimintakäsitteen nimitystä, kuten esimerkissä 298. Kolmessa tapauksessa välissä on käsitteen nimitykseen kuulumattomia sanoja.

**Koetinedokas *нумь*.** Konkordanssihaku hakuehdolla *нумем* (avulla) tuotti 16 osua, joista 11:ssä koetinedokas toimii toimintakäsitteen ja keinon välisen funktiosuhteen koettimena:

- (299) Заделывать пробоину путем постановки мягкого пластыря снаружи – более надежный способ, который удается осуществить почти всегда. (Bezop: 71)

Esimerkissä 299 koetinedokas toimii kahden toimintakäsitteen välisen funktiosuhteen ilmaisimena. Vuotomaton asettaminen reikään (постановка пластыря) on keino, jolla toteutetaan toiminta eli reiän tukkiminen (заделывать пробоину). Voidaan ajatella, että virkkeessä on myös toiminnan ja siihen käytettävän välineen eli vuotomaton (пластырь) välinen suhde. Ilmaus *нумем* on melko produktiivinen ja osumista noin kaksi kolmasosaa on valideja, joten koetin voidaan hyväksyä vahvojen keinon koettimien listaan.

**Agenttiivisuhde ja objektisuhde.** Artikkelilla *Eskos* tehty koettimien poiminta tuotti kaksi mahdollista venäläistä toimijan koetinta. Nämä ovat verbien *осуществлять* (toeuttaa) ja *обеспечивать* (taata, turvata) partisiippi muodot *осуществляющий* ja *обеспечивающий*.

- (300) - - обеспечивает доступ к информации для местных государственных органов, осуществляющих надзор за судоходством (МО, ФПС, ГТК), - - (Eskos: 18)

- (301) - - предоставления информации государственным структурам, обеспечивающим контроль над судоходством, - - (Eskos: 16)

Esimerkissä 300 koetinedokas *осуществляющий* osoittaa, että kollektiivinen toimija paikallinen valtiollinen elin (местный государственный орган) on tekijänä toiminnalle alusliikenteen valvonta (надзор за судоходством). Sulkeissa on lueteltu niiden yksilökäsitteiden nimitykset, jotka kuuluvat yläkäsitteen paikallinen valtiollinen elin (местный государственный орган) alaan. Myös esimerkissä 301 on kollektiivinen toimija, valtion viranomainen (государственная структура), joka on tekijänä toiminnalle alusliikenteen valvonta (контроль над судоходством). Näiden esimerkkien perusteella



partisiippimuodot *осуществляющий* ja *обеспечивающий* toimivat koettimina samalla tavoin.

Konkordanssihaku aineistosta *Bezop* hakuehdolla *осуществляющ\** ei tuottanut yhtään osumaa. Hakuehto *обеспечивающ\** tuotti yhdeksän osumaa, mutta vain yksi on kelvollinen tapaus koetinehdokkaan toiminnasta toimintakäsitteen toimijan koettimena. Osumien vähäisyys osoittaa, että hakuehto saattaa olla liian tarkka. Laajennettu hakuehto *осуществля\** tuottikin 17 osumaa, joista viidessä on toiminta ja toiminnan tekijä.

- (302) Государственный контроль в России осуществляет Главная государственная морская инспекция (Госморинспекция) непосредственно и через службы капитанов портов. (Bezop: 35)
- (303) - - и морская стихия – среда, в которой морской транспорт осуществляет свою деятельность - - (Bezop: 39)
- (304) Ведомственный контроль осуществляется через службы мореплавания парокhodств. (Bezop: 35)

Esimerkissä 302 toiminnan nimitys *государственный контроль* (valtiollinen valvonta) on tekstissä ennen koetinehdokasta ja toiminnan tekijä *Главная государственная морская инспекция* (Valtiollinen merenkulun tarkastusvirasto) koetinehdokkaan jälkeen. Muissakin kelvollisissa osumissa tekijä on ilmaistu tekstissä koettimen jälkeen. Esimerkissä 303 toiminnan tekijä ei ole ihminen tai ihmisryhmä, eikä esimerkkiä tästä syystä ole laskettu kelvolliseksi osumaksi. Konkordanssihaku hakuehdolla *осуществля\** tuotti aineistosta *Bezop* kaksi osumaa, joissa elollisen tekijän sijasta kielipillisenä subjektina on toiminta. Esimerkissä 303 subjektina on merikuljetukset (морской транспорт). Esimerkissä 304 tekijää ei mainita suoraan, vaan toiminta tapahtuu varustamojen merenkulkuosastojen (службы мореплавания парокhodств) kautta.

Tarkennetulla hakuehdolla *обеспечива\** konkordanssihaku tuotti 20 osumaa, joista neljässä on elollinen toimija, mutta vain kahdessa osumassa toimija on toiminnan suorittaja:

- (305) Любое вновь построенное судно будет обладать уровнем надежности, по крайней мере не превышающим технический уровень и культуру производства завода-строителя и поставщиков, обеспечивающих поставку комплектующего оборудования. (Bezop: 42)

Esimerkissä 305 toimittaja (поставщик) vastaa lisälaitetoimituksista (поставка комплектующего оборудования). Muissa osumissa ei ole agenttiivisuhdetta lainkaan.

Kumpikaan toiminnan ja sen tekijän välisen suhteen koetinehdokas ei osoittautunut kelvolliseksi koettimeksi osumien vähäisyyden vuoksi. Parempaa tulosta ei tuonut myöskään toiminnan ja sen kohteen välisten koettimien haku, sillä konkordanssihaku aineistosta *Eskos* ei tuottanut yhtään toiminnan kohdetta osoittavaa koetinehdokasta ja manuaalinen koetinhaku aineistosta *Bezop* tuotti vain yksittäisiä tapauksia koetinehdokaista, jotka eivät konkordanssihaulla osoittautuneet produktiivisiksi.

**Yhteenveto.** Keino- ja menetelmäsuhteen substantiivikoettimet *средство*, *способ* ja *метод* ovat tutkimukseni aineistossa kohtalaisen produktiivisia. Ne kaikki tuottavat kuitenkin paljon roskaa laajalla hakuehdolla. Tarkkuutta on mahdollista parantaa hakuehtoa tarkentamalla tai liittämällä siihen tunnettu termi. Paras tarkkuus saadaan väli-

ne- ja keinosuhteen koettimilla *путем, с помощью* ja *при помощи*, jotka käyvät hakuehdoksi sellaisenaan. Välinesuhteen venäläiset verbikoettimet ovat tutkimukseni aineistossa verbit *применять, использовать, служить* ja *предназначить*, joiden kaikkien tarkkuus nousee noin 70 prosenttiin hakuehtoa muokkaamalla. Verbikoettimista produktiivisin on *применять* (*применяют*). Näihin verbeihin liittyy koettimen roolissa usein prepositio *для*, mutta se ei välttämättä liity verbiin välittömästi. Jos hakuehtoa halutaan tarkentaa preposition avulla, menetetään valideja osumia kaikilla verbeillä. Mutta silloin kun prepositio liittyy välittömästi verbiin, toimintakäsitteen nimitys seuraa välittömästi prepositiota.

## 7.6 Terminologisten koettimien arviointi käsitetiedon poimintamenetelmänä

Jaksossa 7.2.1 mainitsin, että koetintutkimusten tavoitteena on tunnistaa ne kielelliset ilmaukset ja merkit, joiden avulla tekstistä voidaan paikallistaa runsaasti käsitetietoa sisältävät kontekstit (knowledge-rich areas of text) (Bowker 1996b: 35). Koetinpoimintakokeeni tavoitteena oli koota luettelo suomen- ja venäjänkielisistä koettimista, joita voidaan käyttää käsitetiedon poimintaan minkä tahansa erikoisalan sähköisestä korpusesta. Tavoitteena oli myös selvittää, minkälaisen käsitetiedon yhteydessä koettimet esiintyivät. Tässä jaksossa esitän tutkimukseni koettimia koskevan osan tulokset.

Tutkimukseni vahvistaa, että terminologisten koettimien avulla voidaan paikallistaa käsitetietoa sisältäviä konteksteja. Tunnetuilla termeillä tehty konkordanssihakua ja manuaalinen koetinpoiminta tuottivat suomalaisia ja venäläisiä koettimia, joilla erikoisalan teksteistä voi hakea käsitteen nimityksiä, määritelmiä ja vieruskäsitteitä sekä geneerisiä suhteita, koostumussuhteita ja funktiosuhteita. Terminologiset koettimet ovat hyvä apu myös silloin, kun tavoitteena on paikallistaa synonymiaa.

Kuten totesin jaksossa 4.2, koetin voi olla sana, sanaliitto, lyhenne, välimerkki tai painotekninen keino. Eniten osumia tuottavat välimerkkikoettimet, mutta niiden tarkkuus on huono. Koettimina toimivatkin paremmin sanat, ja nämä ovat sanaluokaltaan tavallisesti verbejä tai substantiiveja. Joukossa on sellaisia verbejä ja substantiiveja, joiden tuottamissa osumissa on lähes aina terminologisesti kiinnostavaa tietoa. Nämä koettimet eivät kuitenkaan ole yleensä erityisen produktiivisia, joten niiden tehokas käyttö vaatii suuren aineiston.

Koettimien käyttöön liittyy ongelmia, joista osa johtuu kielen rakenteesta ja osa muista tekijöistä, jotka eivät riipu kielestä. Konkordanssihaun hakuehtoa on vaikea tarkentaa niin, ettei hakuehto olisi liian laaja eikä liian suppea. Liian laaja hakuehto tuottaa suuren määrän kelpaamattomia osumia kun taas liian suppea hakuehto jättää valideja osumia osuimalistan ulkopuolelle. Konkordanssihaun hakuehto olisikin muotoiltava mahdollisimman tarkaksi, jotta osuimalista sisältäisi mahdollisimman paljon terminologista tietoa ilman ylenmääräistä manuaalista osuimien karsintaa. Yksi mahdollisuus on käyttää tarkennettua hakua, kuten Christensen (2000: 243–249) suosittelee. Hänen tutkimuksensa aineisto on tanskankielistä, jolloin prepositioilla voidaan tarkentaa hakua. Venäjän kielessä on sama mahdollisuus, mutta niin sanottu vapaa sanajärjestys ja verbi- rektio heikentävät keinojen käyttökelpoisuutta. Koettimena toimivan verbin ja preposition välissä voi nimittäin olla sanoja, ja sama prepositio liittyy verbiin yleensä myös silloin, kun se ei toimi koettimena. Sekä suomessa että venäjässä nominien sijapäätteet ja verbi- en persoonapäätteet vaikeuttavat haun tarkentamista. Toinen mahdollisuus on korvata taivutusmuodot katkaisumerkillä ja ottaa hakuehtoon vain taipumaton vartalonosa. Tällöin osuimien määrä nousee kuitenkin nopeasti suureksi. Näin käy esimerkiksi verbeillä *sisältyä* ja *sisältää*, jos hakuehdoksi valitaan *sisält\**.

Venäjässä substantiivikoettimien hakuehdon tarkentaminen on vaikeampaa kuin verbien hakuehdon tarkentaminen, sillä verbit ovat koettimena toimiessaan yleensä yksikön tai monikon 3. persoonan preesensmuodossa, mutta substantiivikoettimet voivat saada miltei minkä tahansa sijapäätteen. Esimerkiksi geneerisen suhteen substantiivikoetin *форма* toimii koettimena neljässä tapauksessa, joissa kaikissa se on eri taivutusmuodossa. Hakuehtoon olisi valittava oikeat taivutusmuodot, sillä koettimien käytön tehokkuus riippuu hakuehtojen muotoilun onnistumisesta konkordanssihaussa. Tarkka hakuehto on kuitenkin kömpelö, sillä sen pituus venyy helposti kymmeniin merkkeihin. Jos taas haku tehdään vain joillakin taivutetuilla muodoilla, saattaa tuloksesta jäädä pois valideja osumia.

Eräs keino vähentää roskaosumien määrää on lisätä hakuehtoon termi. Tämä keino on mahdollinen vain siinä tapauksessa, että ainakin osa tarkasteltavan alan termeistä tunnetaan. Välimerkit (sulkeet, kaksoispiste) ja konjunktiot (*sekä, ja*) ovat selkein esimerkki koettimista, joiden käyttö on mielekästä vain yhdessä tunnetun termin kanssa. Myös osa verbeistä ja substantiiveista vaatii seurakseen termin. Tällaisia ovat esimerkiksi suomalaiset *olla* ja *tarkoitus* sekä venäläiset *являться, бывать, форма, часть* ja *средство*. Joidenkin koetinehdokkaiden tarkkuus ei parane edes tunnetun termin avulla, koska niiden ja termin välissä on sanoja. Tällaisia koettimia ovat *osa, syy, tulos* ja *npучина*. Jotkin koettimet toimivat hyvin ja tarkasti ilman termiäkin. Tällaisia ovat esimerkiksi verbit *tarkoittaa, koostua, aiheuttaa, johtua, käyttää, представлять собой, состоять из, включать, применять, вызвать* sekä substantiivit *käsite, aiheuttaja, tehtävä, keino, компонент, помощь* ja lyhenne *ns*.

Kielestä riippumatta koettimet saattavat olla alakohtaisia ja myös kirjoittajakohtaisia, mistä tutkimukseni aineistossa on viitteitä. Suomenkielisistä koetinehdokkaista verbit *koostua, käsittää* ja *osallistua* sekä substantiivit *tyyppi* ja *seuraus* ovat mahdollisesti kirjoittajakohtaisia, koska ne esiintyvät koettimen roolissa vain yhdessä tai kahdessa asiakirjassa. Suomen- ja varsinkin venäjänkielinen aineistoni on kuitenkin liian pieni, jotta kirjoittajakohtaisuudesta voisi tehdä päteviä päätelmiä. Myös tekstilaji näyttäisi selittävän koettimien käyttöä, sillä venäjänkieliseen aineistoon valitussa oppikirjassa koettimia esiintyy enemmän kuin aineistoon kuuluvassa artikkelissa.

Ongelma on myös se, että sama koetin voi toimia monen erilaisen käsitetiedon ilmaisimena (vrt. Soininen 1999a: 56). Esimerkiksi englannin koettimia tutkineen Grinsteadin (2000: 47) mukaan verbikoetin *describe\* as* yhdistyy sekä funktio- että geneeriseen suhteeseen. Omassa tutkimuksessani muun muassa sulkeet on yleiskoetin, ja esimerkiksi määritelmien haku sen avulla on ongelmallista, koska se tuottaa määritelmien lisäksi myös esimerkiksi synonyymeja ja vastineita. Haluttu terminologinen tieto on siksi haettava pitkästä osumalistasta manuaalisesti. Jotkin koettimet tuottavat sekä geneerisen että koostumussuhteen osumia. Osumalistan suppean kontekstin perusteella ei aina pysty sanomaan, kumpi suhde käsitteiden välillä on. Tällaisia koettimia ovat esimerkiksi verbit *sisältää, sisältyä* ja *jakaa*, jotka ovat ennen kaikkea ylä- ja alakäsitteen välisen suhteen koettimia. Hierarkkisten suhteiden koettimet eivät aina toimi toivotulla tavalla, sillä ylä- ja alakäsitteistä kertoessaan kirjoittajat eivät välttämättä käytä lähimpien ylä- ja alakäsitteiden nimityksiä. Kirjoittajat saattavat myös käyttää elliptisiä ilmauksia, jotka on kontekstistaan irrotettuina mahdollista tulkita virheellisesti yläkäsitteen nimityksiksi. Esimerkiksi Sisseck (2005) piirsi tanskankielisestä ravintoainekorpukseltaan koettimien avulla käsitejärjestelmän ja vertasi sitä aiemmin manuaalisesti rakennettuun ontologiaan. Hän havaitsi, että koettimien avulla tuotetusta käsitejärjestelmästä puuttui joitakin välitason käsitteitä, sillä kirjoittaja saattoi esimerkiksi käyttää elliptistä ilmausta *kilder* (lähteet) vaikka viittasi käsitteeseen *A-vitaminkilder* (A-vitamiinin lähteet) (Sisseck 2005: 390).

Luvun 7 alussa tarkastelin suomen ja venäjän koettimia kielittäin sen mukaan, min-  
kälaista terminologista tietoa ne tuottavat. Seuraavaksi teen kummankin kielen koetti-  
mista yhteenvedon sen mukaan, mitä sanaluokkaa tai muuta kieliopillista luokkaa ne  
edustavat. Aloitan yhteenvedon verbikoettimista, koska ne ovat tutkimukseni tuottamis-  
ta suomen ja venäjän koettimista vahvimpia.

### 7.6.1 Verbit koettimina

Aiemmissa koetintutkimuksissa on selvitetty erityisesti verbien toimintaa koettimina. Esimerkiksi Christensen (2000) tutki tanskan kielen verbikoettimia ja totesi niiden il-  
maisevan useimmiten geneeristä suhdetta, koostumussuhdetta tai vieruskäsitteitä. Sak-  
sankielisellä tietosanakirja-aineistolla koettimia tutkinut Nuopponen (1993) totesi, että  
geneeristä suhdetta osoitetaan tietyillä verbeillä (zum ... zählen, umfasst, wird eingeteilt  
in: - -, gliedert sich in, gehören zur, sind). Verbikoettimet osoittautuivat parhaiksi myös  
omassa tutkimuksessani. Niitä on lukumääräisesti paljon, ne ovat produktiivisia eivätkä  
tuota ylen määrin roskaosumia. Poikkeuksiakin on, sillä varsinkin geneerisen suhteen  
suomalainen verbikoetin *on* tuottaa runsaasti roskaa ilman tunnetun termin tukea. Verbi  
tuottaa runsaasti osumia missä tahansa tekstissä, mutta kaikki väitelauseet, joissa *olla*-  
verbi on joko muodossa *on* tai *ovat*, eivät sisällä termejä tai käsitetietoa. Onkin mielen-  
kiintoista havaita, että tutkimuksessani *olla*-verbi ei ilman tunnetun termin tukea ole  
koetin suomenkielisessä aineistossa (ks. jakso 7.4.4), vaikka englanninkielisillä aineis-  
toilla tehtyjen tutkimusten mukaan *to be* on vahva määritelmän (ks. Pearson 1998: 144–  
151) ja geneerisen suhteen koetin (ks. jakso 7.4.4). Grinstead (2000: 39) toteaa saman  
asian sanomalla, että koettimen *is* avulla päästään määritelmän jäljille. Jaksoissa 7.4.3 ja  
7.4.4 totesin kuitenkin, että *olla*-verbin muodot *on* ja *ovat* kannattaa ottaa määritelmän  
tai geneerisen suhteen koettimien listaan, mutta vain siinä tapauksessa, että ne voidaan  
konkordanssihaussa yhdistää tunnettuun termiin, sillä ilman termiä ne tuottavat ylen  
määrin roskaa. Venäjänkielisessä aineistossa eksistentiaalinen, olemassa olemista mer-  
kitsevä *есть*-verbi ei toimi koettimena, sillä sitä käytetään vain harvoin kopulana liit-  
tämään toisiinsa subjekti ja predikaatti. Yleensä kopula jätetään ilmaisematta. Sen sijaan  
venäjän kielessä on lukuisia muita verbejä, joita voidaan käyttää lauseessa kopulan si-  
jasta. Tällainen on esimerkiksi verbi *являться* (*явиться*), joka toimii kuitenkin myös  
koostumussuhteen koettimena. Osittain tästä syystä tämä verbi tuottaa runsaasti roskaa,  
jos vain validit geneerisen suhteen tapaukset hyväksytään.

Käsitteen nimityksen koettimina verbit eivät ole substantiivien veroisia. Tutkimuk-  
seni tuotti vain yhden suomalaisen ja kaksi venäläistä käsitteen nimityksen verbikoetin-  
ta, ja nämäkään eivät olleet erityisen produktiivisia. Verbi *kutsua* on kuitenkin käyttö-  
kelpoinen koetin, koska kaikki sen tuottamat osumat ovat valideja. Myös venäläisten  
koettimien *называется* ja *называемый* tarkkuus on korkea. Kaikille käsitteen nimityk-  
sen verbikoettimille on yhteistä, että termi on lähes aina välittömästi koettimen jälkeen.  
Vastaavan englanninkielisen verbimuodon *is/are called* on todettu toimivan koettimena,  
mutta ainakin Pearsonin (1996: 823) tutkimuksessa se on luokiteltu määritelmän koet-  
timeksi.

Englanninkielisillä aineistoilla tehtyjen tutkimusten mukaan verbit voivat toimia  
myös synonymian koettimina, sillä ilmaukset *also known as* ja *refer\* to as* ovat osoit-  
tautuneet synonymian koettimiksi (Bowker & Pearson 2002: 219; Grinstead 2000: 42;  
Pearson 1998: 183; Kavanagh 1995: 26). Tutkimukseni suomen- ja venäjänkielisten  
koettimien luettelossa ei kuitenkaan ole yhtään synonymian verbikoetinta.

Koetinhaku suomalaisesta aineistosta tuotti kolme määritelmän verbikoetinta. Par-  
haaksi osoittautui verbi *tarkoittaa*, joka tuotti runsaasti valideja osumia ja vähän roskaa.

Se ei myöskään näytä olevan kirjoittajakohtainen. Tulos oli odotettavissa, sillä verbin englanninkielinen vastine *means* on todettu määritelmän koettimeksi ainakin Sierran ja McNaught'n (2000: 10) tutkimuksessa. Myös verbi *määritellä* on tutkimukseni aineistossa kelvollinen määritelmän koetin. Sen englanninkielinen vastine *define* toimii koettimena englanninkielisessä aineistossa (Pearson 1998: 140). Tosin kumpikin suomalainen määritelmän verbikoetin tuotti muodollisten sisältö- tai joukkomääritelmien sijasta vajaita määritelmiä, joissa mainitaan jokin käsitepiirre mutta yläkäsite vain harvoin. Silloin kun yläkäsite on mainittu, se on yleinen tieteen termi (prosessi, toiminta tai kokonaisuus). *Määritellä*-verbin venäläinen vastine *определять* tuotti runsaasti osumia, mutta validien osumien määrä jäi alle raja-arvoksi asetetun kolmen osuman, joten verbi ei ole kelvollinen koetin. Sen sijaan verbit *пониматься*, *понимать* ja *называться* ovat produktiivisia muodollisen määritelmän venäläisiä koettimia. Verbin *называться* englanninkielinen vastine *called* toimii määritelmän koettimena englannin kielessä (Sierra & McNaught 2000:10), mutta tutkimukseni ei tuottanut sille suomalaista koetinvastinetta, kuten ei myöskään englannin koettimelle *refers to*, joka on Sierran ja McNaught'n mukaan (2000: 10) määritelmän koetin. Koetinhaku tuotti aineistosta yhden erityisen vahvan venäläisen verbikoettimen. Se on verbi *представлять собой*, joka tuotti vähän roskaa ja runsaasti sisältö- ja määritelmän muotovaatimukset täyttäviä määritelmiä.

Toisin kuin englanninkielisillä aineistoilla tehdyt tutkimukset (esim. Grinsted 2000: 40–42; Pearson 1996: 818–823), antoivat odottaa, geneerisen käsitesuhteen verbikoettimia löytyi tämän tutkimuksen aineistoista lopulta vähän. Tosin olen kelpuuttanut geneerisen suhteen koettimiksi vain ne verbit, jotka ilmaisevat ensisijaisesti geneeristä suhdetta eivätkä pikemmin esimerkiksi koostumussuhdetta. Ainoa selkeästi geneerisen suhteen suomalainen verbikoetin on *luokitella*-verbi, mutta tutkimukseni perusteella sekään ei ansaitse vahvan koettimen statusta. Verbi *jakaa* toimii useammin koostumussuhteen kuin geneerisen suhteen koettimena, eikä tästä syystä ole vahva geneerisen suhteen koettimena.

Englannin verbikoettimen *include* tavoin suomalaiset vastineet *sisältyä* ja *sisältää* toimivat yleisesti ylä- ja alakäsitteen koettimena, mutta toisin kuin *include* suomalaiset vastineet toimivat pikemmin koostumussuhteen kuin geneerisen suhteen koettimena (ks. Bowker & Pearson 2002: 219; Grinsted 2000: 40–42; Kavanagh 1995: 25, 27). Koettimena niiden toiminta muistuttaakin pikemmin englannin verbiä *contain*, joka on leimallisesti koostumussuhteen koetin (Bowker & Pearson 2002: 219; Grinsted 2000: 43; Kavanagh 1995: 27). Ilman suomalaisia vastineita jäivät englannin geneerisen suhteen verbikoettimet *comprise*, *describe*, *consist(s) of*, *is/are defined as*, *is/are called/known as*, *denote(s)*, *designate(s)* (Grinsted 2000: 40–42; Pearson 1996: 818–823). Venäläisten geneerisen käsitesuhteen verbikoettimien listaan pääsivät verbit *относиться* ja *относиться*.

Koostumussuhteen koettimet eivät joidenkin tutkimusten mukaan näyttäisi olevan teksteissä yhtä yleisiä kuin geneerisen suhteen koettimet. Esimerkiksi oikeustieteen englannin korpusta tutkineet Tariq ym. (2003: 5) toteavat, että heidän aineistossaan koostumussuhteen koettimet eivät olleet erityisen produktiivisia, sillä aineistona olleista 1200 virkkeestä vain 60:ssa oli mahdollinen koostumussuhteen koetin ja niistäkin vain 40 prosenttia todella ilmaisi koostumussuhdetta. Tutkimukseni venäjänkielisessä aineistossa geneerisen suhteen koettimet tuottivatkin huomattavasti enemmän valideja osumia kuin koostumussuhteen koettimet, mutta toisaalta suomenkielisessä aineistossa koostumussuhteen koettimet tuottivat jopa enemmän valideja osumia kuin geneerisen suhteen koettimet. Suomenkielisestä aineistosta nousivat esiin vahvat koostumussuhteen verbikoettimet *koostua*, *jakaa* ja *käsittää*, jotka ovat produktiivisia, tuottavat vähän roskaa ja liittyvät kiinteästi ainakin yhteen termiin. Myös *koostua*-koettimen venäläinen vastine



*состоять из* on produktiivinen, tuottaa vähän roskaa ja esiintyy yhdessä termin kanssa. Verbit *koostua* ja *käsittää* saattavat kuitenkin olla kirjoittajakohtaisia. Verbikoetin *koostua* on aineistossani puhtaasti koostumussuhteen koetin, toisin kuin sen englanninkielinen vastine *consist(s) of*, joka on Pearsonin (1996: 822–823, 1998: 144–151) mukaan myös geneerisen suhteen koetin. Koetinverbi *jakaa* poikkeaa muista koostumussuhteen verbikoettimista siten, että tutkimukseni aineistossa siihen liittyy kolmessa tapauksessa neljästä modaaliapuverbi *voida*. Muista koostumussuhteen verbikoettimista verbit *muodostaa* ja *kuulua* ovat produktiivisia, mutta tuottavat myös roskaa samoin kuin verbit *sisältyä* ja *sisältää*, jotka saattavat lisäksi olla kirjoittajakohtaisia. *Sisältää*-koettimen venäläinen vastine *включать в себя* on sekin vahva koetin, sillä tarkennetulla hakuehdolla se ei tuottanut tutkimukseni aineistossa ollenkaan roskaa. Se ei ole kuitenkaan yhtä produktiivinen kuin *состоять из*. Verbikoetin *muodostaa* vaatii koettimena yleensä vahvistukseksi substantiivikoettimien *osa*, *kokonaisuus*, *osajoukko* tai *käsite*. Englannin ja venäjän verbien *consist* ja *состоять из* mahdollinen vastineverbi *muodostua* ei tutkimukseni aineistossa ollut produktiivinen. Sen toimintaa tulisikin testata isommalla aineistolla. Suomalaisille ja venäläisille koostumussuhteen verbikoettimille on yhteistä, että yläkäsitteen nimitys on ennen koetinta ja alakäsitteen nimitys tai nimitykset sen jälkeen. Vain verbin *muodostaa* yhteydessä järjestys on useimmiten päinvastainen eli alakäsitteen nimitys on ennen koetinta ja yläkäsitteen nimitys koettimen jälkeen.

Syyn ja seurauksen sekä tarkoituksen ja tuloksen koettimista parhaita ovat verbikoettimet, sillä ne ovat produktiivisia, tuottavat tarkennetuilla hauilla vähän roskaa ja niiden valikoima on laaja. Verbi *aiheuttaa* osoittautui hyväksi syyn ja seurauksen välisen käsitesuhteen koettimeksi, sillä se oli sekä erittäin produktiivinen että tarkka. Kaikissa osumissa on syyn ja seurauksen välinen käsitesuhde, ja 83 prosentissa osumista ainakin toinen käsitteistä kuuluu merenkulun erikoisalalle. Myös venäjän aspektiparilliset verbit *привести* ja *приводить* sekä *вызвать* ja *вызывать*, jotka voivat kaikki olla *aiheuttaa*-verbin vastineita, ovat vahvoja syyn ja seurauksen koettimia. Tulos ei ole yllättävä, sillä englanninkielisillä aineistoilla tehdyissä tutkimuksissa verbi *cause* (hakuehdolla *caus\**) on osoittautunut vahvaksi syyn koettimeksi (Bowker & Pearson 2002: 219; Feliu & Cabré 2002: 6; Grinstead 2000: 45–47). Toinen vahva suomalainen syyn ja seurauksen verbikoetin on verbi *aiheutua*, joka käyttäytyy koettimena hyvin samalla tavoin kuin verbi *aiheuttaa*. Kummankin koettimen yhteydessä syykäsitteen nimitys on tekstissä välittömästi ennen koetinta ja seurauskäsitteen nimitys välittömästi koettimen jälkeen suurimmassa osassa valideista osumista. Myös verbin *aiheutua* englanninkielinen vastine *result(s) from* on todettu hyväksi syyn ja seurauksen koettimeksi ainakin Bowkerin ja Pearsonin (2002: 219) sekä Grinsteadin (2000: 45–47) tutkimuksissa. Verbin *result(s) from* toinen mahdollinen vastine *johtua* on sekin hyvä koetin, vaikkei yhtä produktiivinen kuin *aiheutua*. Muista toiminnan tarkoituksen tai tuloksen sekä tapahtuman syyn tai seurauksen verbikoettimista verbi *johtaa* osoittautui vahvaksi syyn ja seurauksen koettimeksi. Tämäkään tulos ei yllätä, sillä ainakin Bowker ja Pearson (2002: 219) ja Grinstead (2000: 45–47) ovat valinneet verbin *johtaa* englanninkielisen vastineen *result in* ja Ahmad ja Salway (1997: 107–108) vastineen *lead to* koettimien luetteloon. Muista mahdollisista suomalaisista seurauksen koettimista verbit *seurata* ja *tuottaa* tuottivat runsaasti roskaa ja vähän valideja osumia. Tämä tulos poikkeaa englanninkielisillä aineistoilla saaduista tuloksista, sillä verbin *tuottaa* vastine *produce* (hakuehto *produces/produced by*) on todettu vahvaksi koettimeksi (Bowker & Pearson 2002: 219; Feliu & Cabré 2002: 6).

Suomalaisessa aineistossa verbit toimivat substantiiveja paremmin myös kausaalisuhteeseen läheisesti liittyvän vuorovaikutussuhteen koettimina. Vuorovaikutussuhteen verbikoettimet tuottavat paljon osumia, mutta ilman tunnettua termiä ne eivät ole



erityisen tarkkoja. Jos vaikutuksen kohdetta nimeävä termi on tunnettu, koettimilla on mahdollista tehdä täsmähakuja, sillä *vaikuttaa*-verbin yhteydessä kohdetta nimeävä termi on yleensä välittömästi ennen koetinta ja verbien *parantaa* ja *vähentää* yhteydessä koettimen jälkeen. Venäläisessäkin aineistossa verbit ovat substantiiveja tarkempia vuorovaikutussuhteen koettimina lukuun ottamatta substantiivia *влияние*, joka on sekä produktiivinen että tarkka. Tarkin venäläinen vuorovaikutussuhteen koetin on kuitenkin ilmaus *за счет*, joka kuuluu vahvojen koettimien joukkoon. Venäläisille vuorovaikutussuhteen koettimille on yhteistä, että vaikuttavaa tekijää nimeävä termi on tavallisesti välittömästi koettimen jälkeen.

Tutkimukseni tuotti vain kaksi vahvaa ja kaksi hyvää suomalaista toimintasuhteen koetinta. Niistä kaksi on verbejä ja kaksi substantiivia. Verbeistä *käyttää* toimii menetelmäsuhteen ja *osallistua* agentiivisuhteen koettimina. Suomalaisten välinesuhteen koettimien joukossa ei ole verbikoettimia, vaikka englanninkielisillä aineistoilla tehdyissä tutkimuksissa nimenomaan verbit (*used for*, *used to*, *employed to*, *designed to*) toimivat välinesuhteen koettimina (Bowker & Pearson 2002: 219; Grinstead 2000: 45–47). Mahdollinen selitys on tarkastelun näkökulmaero. Englannin koettimien yhteydessä huomion kohteena on väline, jonka käyttötarkoitusta partisiippimuodossa oleva koetin ilmaisee, kun taas omassa tutkimuksessani huomio kiinnittyy toimintaan, jonka suorittamisen välinettä koetin osoittaa. Tutkimukseni tuotti runsaasti venäläisiä toimintasuhteen koettimia, joista noin puolet on verbejä ja noin puolet substantiiveja. Produktiivisuudessa ja tarkkuudessa ei ole huomattavaa eroa verbien ja substantiivien välillä, mutta osa substantiivikoettimista vaatii hakuehtoon seurakseen tunnetun termin roskan välttämiseksi. Venäläisessä aineistossa verbit toimivat erityisesti välinesuhteen koettimina toisin kuin suomalaisessa aineistossa. Tämä tulos on samansuuntainen kuin englanninkielisillä aineistoilla. Vahvoja välinesuhteen koettimia ovat verbit *применять*, *использовать*, *служить* ja *предназначить*, jotka ovat sekä produktiivisia että tarkkoja.

Olen hyväksynyt valideiksi osumiksi myös muita kuin verbien preesensmuotoja, vaikka automaattiset haut perustuvat käsitykseen, että verbimuotoisen käsitesuhdetta osoittavan ilmauksen tulisi olla preesensissä, koska vain silloin suhde on yleispätevä. Tämän käsityksen mukaan modaaliapuverbit eivät ole sallittuja, koska ne rajoittavat lausuman yleispätevyyttä ja osoittavat epävarmuutta. Modaaliapuverbien käyttö teksteissä on kuitenkin yleistä, eivätkä ne välttämättä tarkoita sitä, että kirjoittaja olisi epävarma asiastaan. Kirjoittaja saattaa käyttää modaaliapuverbiä, koska hän haluaa jättää muille mahdollisuuden esimerkiksi määritellä käsite toisin. Modaaliapuverbin käyttö määritelmässä saattaa myös liittyä käsitteen elinkaareen. Käsite on ehkä vasta elinkaarensa alussa eikä sen määritelmä ole vielä vakiintunut, vaan asiantuntijat vasta neuvottelevat siitä.

Aineistossani modaalisuus liittyy erityisesti syyn ja seurauksen välisen suhteen koettiin. Syy ei siis välttämättä johda seuraukseen. Modaalisuutta liittyy myös hierarkkisten ylä- ja alakäsitteiden väliseen suhteeseen, jolloin kirjoittaja on halunnut antaa mahdollisuuden luokitella käsitteet eri tavalla kuin hän on ne luokitellut. Geneerisen suhteen koettimista modaalisuutta liittyy erityisesti suomen verbiin *jakaa* ja venäjän verbiin *отнести*, johon saattaa liittyä predikaatiivinen adverb *можно*.

## 7.6.2 Substantiivit koettimina

Suomenkielisistä käsitettä nimeävän ilmauksen koetinehdokkaista parhaaksi osoittautui tämän tutkimuksen aineistolla substantiivi *käsite*, joka oli sekä produktiivinen että tarkka. Myös sen venäjänkieliset vastineet *понятие* ja *концепция* ansaitsevat koettimen

statuksen. Niistä *понятие* oli produktiivisempi ja tarkempi. Substantiivit *понятие* ja *концепция* saattavat kuitenkin olla kirjoittajakohtaisia, sillä niiden frekvenssi eri asia-kirjoissa sivumääriin suhteutettuna eroaa huomattavasti. Kumpaakin koetinta seuraa välittömästi termi melkein kaikissa valideissa osumissa, joten osuimien käsittely on nopeaa. Englanninkielisillä aineistoilla tehdyissä tutkimuksissa substantiivi *the term* on osoittautunut selkeäksi termin koettimeksi (Pearson 1998: 133; Sierra & McNaught 2000: 10). Tässä tutkimuksessa sen suomenkielinen vastine *termi* ja venäjänkielinen vastine *термин* vahvistavat tämän tuloksen suomen- ja venäjänkielisillä aineistoilla. Myös substantiivi *nimitys* osoittautui kelvolliseksi koettimeksi. Se ei ollut yhtä produktiivinen kuin *termi*, mutta roskan määrä jäi siltäkin alhaiseksi.

Synonymian koettimina substantiivit eivät näytä toimivan, sillä tutkimukseni koetinhaku ei tuottanut yhtään suomalaista tai venäläistä synonymian substantiivikoetinta. Myöskään määritelmän koettimet eivät ole substantiiveja, sillä koetinhaku tuotti vain yhden suomalaisen määritelmän substantiivikoettimen, substantiivin *määritelmä*. Se ei ollut produktiivinen, mutta osoittautui vahvaksi määritelmän koettimeksi, sillä se tuotti erittäin vähän roskaa. Sen sijaan sen venäläinen vastine, substantiivi *определение*, tuotti vain vähän valideja tapauksia.

Substantiivi *tyyppi* tuotti aineistossani yli 10 osumaa, joista liki puolet on valideja geneerisen suhteen konteksteja, mutta koetin tuotti vähän uutta tietoa ja oli kirjoittajakohtainen. Myös sen venäläinen vastine *мун* tuotti vain vähän uutta tietoa. Enemmän valideja osumia tuottikin substantiivi *евд*. Nämä tulokset ovat osittain ristiriidassa englanninkielisillä aineistoilla saatujen tulosten kanssa, sillä substantiivit *a kind of* ja *a type of* on mainittu geneerisen suhteen koettimiksi ainakin kolmessa tutkimuksessa (Ahmad & Rogers 2001: 749; Bowker & Pearson 2002: 219; Feliu & Cabré 2002: 5).

Substantiivit ovat verbejä heikompia koostumussuhteen koettimina varsinkin suomen kielessä. Suomalaisten koostumussuhteen koetinehdokkaiden joukossa oli vain yksi substantiivi, *osa* ja siitä muodostettu yhdyssana *osatekijä*. Substantiivi *osa* tuotti aineistostani runsaasti osumia, mutta niistä vain pieni osa on valideja koostumussuhteen osumia. Aineistossani se toimiikin paremmin vaikutussuhteen kuin koostumussuhteen koettimena, vaikka roskan vähentäminen on tällöinkin ongelmallista. Hakua on vaikea tarkentaa tunnetun termin avulla, koska koetin liittyy harvoin kiinteästi termiin. *Osa*-substantiivin venäläinen vastine *часть* tuotti sekin vähän valideja koostumussuhteen osumia. Tulos on osittain ristiriidassa englanninkielisillä aineistoilla saatujen tulosten kanssa, sillä substantiivia *part* (osa) pidetään leimallisesti koostumussuhteen koettimena (Feliu & Cabré 2002: 7; Grinstead 2000: 43; Kavanagh 1995: 27). Venäläisissä koostumussuhteen substantiivikoettimissa on kuitenkin valinnanvaraa, sillä substantiivin *часть* lisäksi aineistostani nousi viisi venäläistä koostumussuhteen substantiivikoetinta. Parhaaksi niistä osoittautui *состав*. Esimerkiksi Grinstead (2000: 44) sekä Feliu ja Cabré (2002: 7) ovat todenneet, että substantiivit *element* ja *component* ovat englanninkielisiä koostumussuhteen koettimia. Niiden vastineet *элемент* ja *компонент* osoittautuivat koettimiksi venäläisessä aineistossa, mutta kumpikaan niistä ei ole vahva koetin.

Substantiivikoettimet eivät tutkimukseni aineistossa ole aivan verbien veroisia syyn ja seurauksen tai tarkoituksen ja tuloksen koettimia. Suomalaiset substantiivikoettimet eivät ole aivan yhtä produktiivisia tai tarkkoja kuin verbikoettimet, mutta ovat kuitenkin käyttökelpoisia. Suomalaisista syyn ja seurauksen substantiivikoettimista produktiivisin tutkimusaineistossani oli *syy*. Tarkin syyn ja seurauksen substantiivikoetin oli *aiheuttaja*, joka toimii koettimena ilman termiäkin. Toiminnan tarkoituksen ja tuloksen koettimista produktiivisimpia olivat *tavoite* ja *tarkoitus*, joista jälkimmäinen toimii koettimena parhaiten yhdessä tunnetun termin kanssa. Tarkin oli *tehtävä*. Venäläisistä tarkoituksen ja tuloksen koettimista produktiivisin ja tarkin oli *результат*. Tutkimukseni tuotti

kolme muutakin venäläistä kausaalisuhteen substantiivikoetinta, mutta niiden tarkkuus on alhainen eivätkä ne ole erityisen produktiivisia. Tutkimukseni tuotti myös runsaasti suomalaisia ja venäläisiä vuorovaikutussuhteen substantiivikoettimia, jotka eivät kuitenkaan olleet aivan yhtä produktiivisia ja tarkkoja kuin saman suhteen verbikoettimet. Suomalaisen *tekijä*-koettimen tarkkuutta on kuitenkin mahdollista parantaa lisäämällä hakuehtoon määrite ja *vaikutus*-koettimen tarkkuutta lisäämällä ehtoon vaikuttavaa teki-jää tai vaikutuksen kohdetta nimeävä termi.

Tutkimukseni tuotti vain kaksi suomalaista toimintasuhteen substantiivikoetinta. Näistä produktiivisempi on väline- tai keinosuhteen koetin *apu*. Sen tarkkuutta on mahdollista parantaa liittämällä hakuehtoon tunnettu termi, sillä välineen nimitys on lähes aina välittömästi ennen koetinta. Keinosuhteen koetin, substantiivi *keino*, ei ollut yhtä produktiivinen, mutta sen tarkkuus on hyvä ilman tunnettua termiäkin. Venäläisiä toimintasuhteen substantiivikoettimia tutkimuksessani on suomalaisia enemmän niin kuin verbikoettimiakin. Venäläisessä aineistossa substantiivit toimivat erityisesti keinosuhteen koettimina. Suomalaisten koettimien tavoin osa venäläisistä substantiivikoettimista tuottaa runsaasti roskaa, ellei hakua ole mahdollista rajata tunnetulla termillä, ja osa toimii hyvin ilman termiäkin. Vähän roskaa tuottavista koettimista *метод* on vahva menetelmän koetin ja *путь* vahva keinon koetin. Myös *помощь* on tarkka tarkennetulla haulla, vaikka osumien määrä on alhainen. Keinon substantiivikoettimet *средство* ja *инструмент* ovat produktiivisia mutta tuottavat paljon roskaa.

### 7.6.3 Muut leksikaaliset koettimet

Suomalaisista lyhenteistä *ns.* osoittautui hyväksi käsitettä nimeävän ilmauksen koettimeksi, sillä se on produktiivinen eikä tuota runsaasti roskaa. Myös lyhenne *mm.* on kelvollinen geneerisen suhteen koetin, joka tuottaa vähän roskaa yhdistettynä *olla*-verbin muotoon *ovat*. Se näyttäisi olevan nimenomaan suomalainen koetin, sillä sen vastinetta ei ole vastaavien englanninkielisten koettimien luettelossa. Lyhennekoettimen vastine toisessa kielessä ei aina olekaan lyhenne, vaan esimerkiksi konjunktio. Esimerkiksi englanninkielisen synonymian koettimen *i.e.* (Bowker & Pearson 2002: 219; Pearson 1998: 176–177) lähivastine suomessa on *eli*. Venäjänkielisistä koetinehdokkaista lyhenne *m.e.* ja sen synonyymi *иными словами* tuottivat kumpikin vain yhden validin osuman. Muutkaan lyhenteet eivät toimineet synonymian koettimina tutkimukseni venäläisessä aineistossa. Sen sijaan lyhenneilmaus *u dp.* näyttäisi toimivan käsitteen nimityksen koettimena, sillä sitä edeltää termien luettelo.

Englanninkielisten geneerisen suhteen koettimien *e.g.* ja *for example* (Grinstead 2000: 40–42; Pearson 1998: 124) suomalainen vastine *esimerkiksi* ei tämän tutkimuksen aineistolla yltänyt vahvojen koettimien joukkoon. Se tuotti runsaasti roskaa ja toimi koettimena vain yhdessä konjunktion *kuten* kanssa, mutta tällöin osumien lukumäärä jäi alhaiseksi. Venäläinen vastine *например* on otettu koettimien listaan, mutta vahvaksi koettimeksi ei sitäkään voi sanoa, koska ilman termiä se tuottaa paljon roskaa. Englannin koettimista jäivät kokonaan vaille suomalaista vastinetta esimerkiksi adverbit *especially*, *chiefly* ja *notably* (ks. Grinstead 2000: 40–42; Kavanagh 1995: 25).

Konjunktio *eli* osoittautui hyväksi synonymian mutta heikoksi määritelmän koettimeksi. *Eli* ei ole synonymian koettimena yhtä produktiivinen kuin sulkeet, eikä sen tarkkuuskaan yllä korkealle, mutta sen käyttökelpoisuutta lisää se, että kaikissa valideissa tapauksissa sen välittömässä läheisyydessä on termi. *Eli*-konjunktio toimii myös määritelmän koettimena, mutta roskaosumien suuri määrä heikentää sen käyttökelpoisuutta tässä tehtävässä. Sen venäjänkieliset vastineet *то есть*, *иными словами* ja *или* eivät tutkimukseni aineistossa osoittautuneet synonymian koettimiksi. Vastaava englan-

ninkielinen koetin lyhenne *i.e.* toimii Sierran ja McNaught'n (2000: 10) mukaan määritelmän koettimena.

Konjunktion ja adjektiivin yhdistelmä *and other* on Grinstedin mukaan (2000: 40) vieruskäsitteiden koetin englannin kielessä. Nuopponen (1993: 107–110) havaitsi, että *ja*-konjunktion saksankielinen vastine *und* sekä lähivastine *ausserdem* (sen lisäksi) toimivat vieruskäsitteiden koettimina. Nämä tulokset saavat tukea tutkimuksestani, sillä konjunktio *ja* toimii vieruskäsitteiden koettimena yhdessä yhdysmerkin kanssa. Myös *ja*-konjunktion venäläinen vastine *u* toimii koettimena, mutta vain yhdessä termin kanssa. Sen sijaan konjunktion ja lyhenteen yhdistelmä *u òp.* toimii tutkimukseni aineistossa vieruskäsitteiden koettimena myös ilman termiä.

Konjunktio *kuten* on produktiivinen geneerisen suhteen koetin. Lisäksi yläkäsitteen nimitys on kaikissa valideissa tapauksissa välittömästi ennen koetinta ja alakäsitteen nimitys välittömästi koettimen jälkeen melkein kaikissa tapauksissa. Koetin *kuten* toimii siis samalla tavalla kuin sen englanninkielinen vastine *such as*, jota edeltää Grinstedin (2000: 40–42) mukaan yläkäsitteen nimitys ja seuraa alakäsitteiden luettelo.

Adjektiivit tai pronominit eivät yleensä toimi koettimina, mutta tähänkin sääntöön on poikkeus. Pronomini *muu* on vahva geneerisen suhteen koetin, joka viihtyy termin seurassa ja tuottaa vähän roskaa yhdistyneenä konjunktioihin *ja*, *tai* tai *myös*. Venäläinen vastine *òpyzoù* on vahva geneerisen suhteen koetin venäläisessä aineistossani. Tämä tulos noudattaa englanninkielisillä aineistoilla saatuja tuloksia, joiden mukaan *and/or other* on geneerisen suhteen koetin (Grinsted 2000: 40–42; Kavanagh 1995: 25).

Englannin kielessä omistajan ilmaisemiseen käytetyt kielelliset keinot pääte *'s*, prepositio *of* ja ilmaus *has a* voivat toimia koostumussuhteen koettimina (Ahmad & Rogers 2001: 749; Bowker & Pearson 2002: 219; Feliu & Cabré 2002: 7; Grinsted 2000: 43; Kavanagh 1995: 27). Suomen kielessä omistajaa ilmaistaan sijamuodoilla ja venäjän kielessä sijamuodoilla ja prepositioilla. Sijapäätteet ovat kuitenkin aivan liian yleisiä, jotta niitä voisi käyttää koettimina. Englannin kielessäkin omistajaa ilmaisevien kielellisten keinojen käyttökelpoisuus koettimena on kyseenalainen, sillä esimerkiksi Grinstedin (2000: 43) aineistossa possessiivinen *'s* oli ongelmallinen koostumussuhteen koettimena, sillä se tuotti erittäin paljon roskaa.

Venäläisestä aineistosta tarkastelin myös joidenkin prepositioiden toimintaa koettimena. Prepositiot ovat kuitenkin yleensä liian epätarkkoja koettimiksi ilman termiä, koska ne voivat toimia monessa eri tehtävässä. Säännöstä on yksi poikkeus, sillä syyn ja seurauksen koetin *u3-3a* toimii koettimena ilman termiäkin. Sen tarkkuus paranee entisestään, jos haku on mahdollista tehdä preposition ja syykäsitettä nimeävän termin yhdistelmällä, sillä termi seuraa välittömästi koetinta lähes kaikissa valideissa tapauksissa.

#### 7.6.4 Ei-leksikaaliset koettimet

Ei-leksikaaliset koettimet toimivat vain yhdessä tunnetun termin kanssa, sillä ilman termiä ne tuottavat ylen määrin roskaa. Sulkeet ovat manuaalisen koetinhaun perusteella hyvin monipuolinen koetin, sillä niitä käytetään teksteissä monenlaisen terminologisesti kiinnostavan tiedon yhteydessä. Käsinpoiminnan perusteella sulkeet voivat olla ainakin käsitettä nimeävän ilmauksen, synonymian, geneerisen suhteen ja määritelmän koetin sekä suomalaisessa että venäläisessä aineistossa. Myös Nuopponen (1993) havaitsi, että hänen tutkimassaan saksankielisessä aineistossa sulkeet toimivat geneerisen käsitesuhteen koettimena. Sulkeiden käyttö koettimena on kuitenkin ongelmallista, sillä ne tuottavat runsaasti osumia, jotka on lajiteltava manuaalisesti. Osumien joukossa on myös paljon sellaisia konteksteja, jotka eivät ole terminologisesti kiinnostavia.

Englanninkielisillä aineistoilla tehdyissä tutkimuksissa sulkeet ovat osoittautuneet synonymian koettimeksi (vrt. Ahmad & Rogers 2001: 749; Pearson 1998: 174). Suomalaisesta aineistosta sulkeet tuottivat kuitenkin ennen kaikkea englanninkielisiä vastineita. Tutkimukseni aineistossa sulkeet toimivatkin ennemmin vastineen kuin synonyymin koettimena. Tosin englanninkielisiä lyhennetermejä voidaan monessa tapauksessa pitää suomalaisina termeinä, koska ne ovat suomenkielisissä merenkulun teksteissä hyvin yleisiä. Venäläisessä aineistossa sulkeet toimivat hyvinkin eri tavalla eri tekstilajeissa. Aikakauslehden artikkelista ne tuottivat termin lyhyitä muotoja ja oppikirjatekstistä termejä, joiden synonyymi tai yläkäsite on mainittu tekstissä ennen sulkeita.

Geneeristen suhteiden haussa sulkeet ovat käyttökelpoinen koetin, jos hakuehdossa on mahdollista käyttää tunnettua termiä. Määritelmän koettimena sulkeet toimivat heikosti, sillä ne tuottivat runsaasti osumia, joissa ei ole edes vajaata määritelmää. Määritelmän koettimena sulkeet toimivatkin parhaiten yhdessä sulkeissa olevan tunnetun termin kanssa, sillä määritelmätieto on yleensä ennen sulkeissa olevaa termiä sekä suomalaisessa että venäläisessä aineistossa. Osumien monilukuisuuden vuoksi sulkeiden käyttö koettimena on yleensäkin mielekäästä vain termin kanssa yhdessä, jolloin osumien määrää voidaan rajoittaa ja manuaalisen työn osuutta vähentää.

Englanninkielisillä aineistoilla tehtyjen tutkimusten tuloksista poiketen (esim. Pearson 1998: 102) pilkku ei osoittautunut synonymian koettimeksi sen paremmin suomenkielisessä kuin venäjänkielisessäkään aineistossani. Myöskään kaksoispiste ei toimi koettimena, vaikka esimerkiksi Sierra ja McNaught (2000: 10) mainitsevat kaksoispisteen termin koettimena ja Nuopposen (1993) tutkimassa saksankielisessä aineistossa kaksoispiste ilmaisi geneeristä suhdetta. Testasin kaksoispistettä geneerisen suhteen koettimena, mutta ilman tunnettua termiä se tuottaa runsaasti roskaosumia, joiden karsinta vaatii osumalistan manuaalista käsittelyä. Tutkimukseni perusteella kaksoispiste toimii kuitenkin koettimena, jos hakuehtoon voi liittää tunnetun termin, sillä kaikissa valideissa tapauksissa välittömästi kaksoispisteen jälkeen on termi.

Pearson (1998: 124) pitää lainausmerkkejä termin koettimena, jota käytetään silloin, kun termi on uusi tai sanalla on yleiskielestä poikkeava merkitys, vaikka sana ei olisi-kaan varsinainen termi. Suomalaisessa aineistossa lainausmerkit oli produktiivinen käsitettä nimeävän ilmauksen koetinehdokas, mutta se tuotti myös runsaasti roskaa (vrt. Pasanen 2005a: 295). Koska kuitenkin noin kolmannes osumista oli valideja, lainausmerkit on hyväksytty koettimien luetteloon.

Ei-leksikaalisiin koettimiin voidaan luokitella myös ajatusviivan ja pronominin *эмо* yhdistelmä. Se on venäläinen määritelmän koetin, jolle ei ole vastinetta englannissa tai suomessa. Koetin ei ollut erityisen produktiivinen, mutta joka toisessa sen tuottamassa osumassa on määritelmä, joka on kaikissa tapauksissa sisältömääritelmä. Lisäksi määriteltävää käsitettä nimeävä termi on kaikissa valideissa osumissa välittömästi ennen koetinta.

Tutkimukseni tulosten perusteella suomalaisissa ja venäläisissä merenkulun turvallisuutta käsittelevissä teksteissä parhaiten koettimina toimivat verbit, jotka tuottavat tietoa erityisesti keskenään kausaalisuhteessa olevista käsitteistä. Eräät substantiivit ovat koettimina lähes verbien veroisia, sillä ne ovat produktiivisia ja riittävän tarkkoja puoli-automattisiin menetelmiin. Kaikkien koettimien käyttökelpoisuus paranee, jos niihin voi konkordanssihaussa yhdistää tunnetun termin.



## 8 JOHTOPÄÄTÖKSET

Tutkimuskohteenani olivat termit, terminpoimintamenetelmät ja terminologiset koettimet, joiden avulla teksteistä voidaan poimia terminologista tietoa. Toivon tutkimuksestani olevan hyötyä ennen kaikkea deskriptiivistä sanastotyötä tekeville terminologeille, kääntäjille ja tulkeille sekä erikoisalan opettajille ja opiskelijoille, jotka poimivat termejä ja käsitetietoa teksteistä. Heille on tarpeen nähdä termit todellisessa käyttökontekstissaan. Tutkimukseni edustaaakin uudempaa terminologian suuntaa, joka keskittyy deskriptiiviseen ja tekstilähtöiseen tutkimukseen. Deskriptiivisen sanastotyön avuksi on kehitetty uusia menetelmiä, joissa aineistona käytetään tekstikorpuksia. Käsitetiedon paikallistamiseksi teksteistä on ensin paikallistettava termit. Silloin käsiteanalyysin lähtökohta onkin termi eikä käsite, niin kuin perinteisessä käsiteanalyysissä. Termien ja käsitetiedon paikallistamiseen on jo olemassa puoliautomaattisia menetelmiä, joiden perustana ovat hakusanojen tunnistamiseen, indeksointiin ja tietopalveluun kehitetyt menetelmät ja esimerkiksi korpuslingvistiikan alalta saadut kokemukset. Vaikka näiden menetelmien avulla teksteistä tunnistetaan termejä ja käsitetietoa ja järjestetään se niin, että se on helposti käytettävissä, niitä ei kuitenkaan ole kehitetty nimenomaan sanastotyön tarpeisiin. Työni tarkoitus onkin luoda pohjaa menetelmien edelleen kehittämiseksi niin, että ne vastaisivat paremmin deskriptiivistä sanastotyötä tekevien terminologien tarpeita. Työni analyysiosia on laaja ja yksityiskohtainen. Toivon, että sellaisena siitä on apua terminologisen tiedon poimintaan tarkoitettujen tietoteknisten työkalujen kehittäjille. Teoreettisesti ja menetelmällisesti työni noudattaa lähinnä pohjoismaista tutkimusperinnettä, jossa korostuu soveltava terminologian tutkimus ja sanastotyö, mutta olen saanut vaikutteita myös sosioterminologiasta, erityisesti venäläisestä kognitiivisesta terminologiasta. Tutkimusmenetelmäni toimivuutta voi testata toteuttamalla vastaavia tutkimuksia muiden erikoisalojen aineistoilla ja muilla kielillä. Tutkimani terminologisen tiedon poiminnan menetelmä antaa rakennusaineita kehysteorian soveltamiseen merenkulun alalla. Suunnitelmissani onkin tutkimuksen suuntaaminen kognitiiviseen terminologiaan ja kehysteorian soveltamiseen merenkulun erikoisalaan.

Työssäni on kaksi kieltä, suomi ja venäjä. Suomen kielen valitsin ennen kaikkea siksi, että se on äidinkieleni ja ajatusten muotoilu tekstiksi on sillä luontevinta. Halusin kirjoittaa työni suomeksi myös siksi, että suomen kielestä on tehty vain vähän laajempaa terminologista tutkimusta. Vieläkin vähemmän on tutkittu merenkulun suomea, jota ei tietääkseni ole aikaisemmin tarkasteltu terminologian näkökulmasta. Ajastamme kertoo sekin, että olen joutunut puolustelemaan kielivalintaani. Mielestäni on kuitenkin tärkeää, että kaikilla erikoisaloilla on mahdollisuus kirjoittaa ja viestiä suomeksi tieteen kielellä ja että erikoisalojen käsitteille on myös suomenkieliset nimitykset. Haluan tarjota terminologiaa opiskeleville halutessaan mahdollisuuden lukea alan tieteellistä kirjallisuutta myös suomeksi, sillä uusien käsitteiden ymmärtäminen on helpompaa, jos niihin voi tutustua omalla äidinkielellään. Venäjän valinta toiseksi kieleksi oli luontevaa, sillä tutkimukseni kuuluu venäjän kääntämisen ja tulkkauksen oppiaineeseen. Venäjän kieli näkyykin vahvasti työni teoreettisessa osassa, sillä olen työssäni tuonut esille venäläistä terminologian tutkimusta, jolla on pitkät perinteet. Laajuudestaan huolimatta se tunnetaan länsimaissa huonosti. Sillä olisi kuitenkin paljon annettavaa muissa maissa tehtävälle tutkimukselle.

Tavoitteeni ei ole ollut suomen ja venäjän kielen vertailu, mutta tutkimukseni tuo esiin joitakin kielten välisiä eroja. Terminpoiminnan analyysi osoittaa, että suomalaisissa ja venäläisissä termeissä on kielistä johtuvia periaatteellisia eroja. Suomenkielisissä termeissä näkyy pyrkimys lyhyteen tai tiiviiseen, sillä tutkimukseni perusteella yli puolet merenkulun suomen termeistä on yksisanaisia. Venäjänkielisissä termeissä näkyy



puolestaan pyrkimys tarkkuuteen, sillä huomattava osa termeistä on sanaliittoja ja yksisanaisten osuus jää neljännekseen.

Tutkimuksen empiiriseen osuuteen kuuluvassa luvussa 6 selvitin, miten hyvin toisaalta suomalaiset ja venäläiset merenkulun ja kieliaineiden tai kääntämisen opiskelijat ja toisaalta eräät kaupalliset terminpoimintaohjelmat selviävät terminpoiminnasta. Opiskelijaryhmien ja terminpoimintaohjelmien tuottamia termiehdokaslistoja vertasin erikoisalan asiantuntijoiden ja terminologien poimimien termien listaan, jota pidin mahdollisimman oikeana termilistana.

Kokeen käsinpoimintaosiossa asetin tavoitteeksi vastata kysymykseen, onko erikoisalan tulevien asiantuntijoiden ja kielten tai kääntämisen tulevien asiantuntijoiden terminvalinnassa eroja, jotka selittyvät erikoisalan tai terminologian tuntemuksella tai sen puutteella. Vertasin suomalaisista merenkulun tai kääntämisen ja venäläisistä merenkulun tai kieliaineiden opiskelijoista koottujen koehenkilöryhmien valitsemien termiehdokkaiden sanalukua, frekvenssiä ja termikaavoja asiantuntijoiden valitsemien termien vastaaviin ominaisuuksiin. Jaoin vertailun kolmeen osaan. Ensimmäisessä osassa tarkastelin termejä, jotka opiskelijat valitsivat, toisessa osassa termejä, joita opiskelijat eivät valinneet, eli niin sanottuja piilotermejä, ja kolmannessa osassa termiehdokkaita, jotka eivät ole termejä vaan niin sanottuja valetermejä.

Koehenkilöiden valitsemien termiehdokkaiden määrä oli keskimäärin hyvin lähellä asiantuntijoiden valitsemien termien määrää, mutta yksilölliset erot valittujen termiehdokkaiden määrissä olivat huomattavasti suuremmat kuin ryhmien väliset erot. Suomalaisen koehenkilöiden ryhmistä yksilölliset erot olivat suuremmat kääntämisen opiskelijoiden ryhmässä kuin merenkulun opiskelijoiden ryhmässä. Termiehdokkaiden lukumäärä vaihteli koehenkilöittäin huomattavasti enemmän kääntämisen opiskelijoiden ryhmässä. Sen sijaan venäläisistä koehenkilöryhmistä vaihtelu oli suurempaa merenkulun opiskelijoiden ryhmässä. Kokeeseen osallistuneet suomalaiset merenkulun opiskelijat olivat kokeen aikana suorittamassa ammattiinsa liittyviä jatko-opintoja. Kaikilla heillä oli merenkulun alan työkokemusta, jonka kautta kertynyt erikoisalatieto ilmeisesti tasoitti yksilöllisiä eroja suomalaisten merenkulun opiskelijoiden ryhmässä.

Kielten sisäiset koehenkilöryhmien väliset erot termien ja termiehdokkaiden pituudessa, frekvenssissä ja termikaavoissa olivat pieniä. Suomalaiset merenkulun opiskelijat valitsivat vähemmän kaksi- ja kolmisanaisia ja pienifrekvenssisiä termejä ja termiehdokkaita kuin kääntämisen opiskelijat. Venäläiset merenkulun opiskelijat valitsivat runsaasti korkeintaan kaksisanaisia valetermejä. Heidän valitsemiensa termiehdokkaiden listassa on myös runsaasti vähintään nelifrekvenssisiä valetermejä. Piilotermien pituudessa ja frekvenssissä oli kielten välisiä eroja. Suomalaiset koehenkilöt jättivät poimimatta yksi- ja kaksisanaisia termejä, kun taas venäläiset koehenkilöt poimivat hyvin yksisanaiset termit mutta jättivät poimimatta toisaalta pienifrekvenssisiä ja toisaalta vähintään kolmisanaisia termejä.

Kummankin suomalaisen koehenkilöryhmän valitsemien termiehdokkaiden yleisimmät termikaavat ovat samat, ja niiden mukaisten termiehdokkaiden suhteelliset osuudet ovat hyvin lähellä toisiaan. Merenkulun opiskelijat poimivat kuitenkin adjektiiveja ja verbejä sekä rinnastuskonjunktion sisältäviä sanaliittoja, joita kääntämisen opiskelijat eivät poimineet. Sen sijaan kääntämisen opiskelijat poimivat mahdollisesti saamansa koulutuksen ohjaamina adjektiivin ja substantiivin sanaliittoja. Myös venäläiset merenkulun opiskelijat poimivat termilistaansa adjektiiveja toisin kuin venäläiset kieliaineiden opiskelijat. Tähän on ilmeisesti vaikuttanut kieliaineiden opiskelijoiden tietämys terminologian teoriasta. Kaikki manuaaliseen terminpoimintaan osallistuneet poimivat termejä käsitelähtöisesti riippumatta koulutuksesta tai erikoisalan tuntemuksesta.

Erilaisilla menetelmillä tuotettuja termiehdokaslistoja vertaamalla pyrin selvittämään, millaisia ovat koetekstin keskeiset termit, jotka eivät tuota vaikeuksia terminpoiminnassa. Termilistoissa ja termiehdokaslistoissa onkin joukko termejä, jotka kaikki terminpoimintaan osallistuneet valitsivat. Nimitän näitä termejä ydintermeiksi. Totesin, että ydintermeillä on suuri frekvenssi toisin kuin termeillä yleensä, sillä suurin osa asiantuntijoiden valitsemista termeistä esiintyy tekstissä vain kerran. Suomalaiset ydintermit ovat suurimmaksi osaksi yksisanaisia yhdyssubstantiiveja, joiden frekvenssi on vähintään kolme. Venäläisistä ydintermeistä suurin osa on kaksisanaisia adjektiivin ja substantiivin sanaliittoja, jotka esiintyvät tekstissä vähintään kolmesti.

Pyrin myös selvittämään, millaiset termit tuottavat vaikeuksia manuaalisessa ja puoliautomaattisessa terminpoiminnassa ja kuinka käyttökelpoisia ovat kaupallisten terminpoimintaohjelmien tuottamat termiehdokaslistat. Koehenkilöiden termiehdokaslistoista puuttuu monisanaisia termejä, joilla esiintyy variaatiota. Termiehdokkaiden pituudessa oli vaihtelua, sillä koehenkilöt eivät olleet yksimielisiä siitä, mistä monisanainen termiehdokas alkaa tai mihin se päättyy. Suomalaisen koehenkilöiden valinnasta puuttuu lisäksi joitakin erikoisalan termejä, joilla on myös yleiskielinen merkitys. Koehenkilöiden poimimat valetermit taas ovat yleiskielen sanoja tai liian pitkiä termeiksi. Usein liian pitkissä termiehdokkaissa on kaksi termiä.

Puoliautomaattiset terminpoimintaohjelmat poimivat hyvin ydintermit, joiden frekvenssi on vähintään kolme, ja suurimman osan termeistä, joiden frekvenssi on vähintään kaksi. Terminpoimintaohjelmien tuottamien termiehdokaslistojen ongelma on valetermien runsaus. Testaamani terminpoimintaohjelmat kelpaavat nykymuodossaan ydintermien hakuun, minkä jälkeen käyttäjä voi päättää, kannattaako teksti valita terminologiseen analyysiin. Tähän riittävät parhaat pisteet saaneet termiehdokkaat, jotka todella ovatkin termejä kaikilla testatuilla ohjelmilla. Roskan määrä on mahdollista säätää pieneksi, jolloin ohjelmat tuottavat lyhyen termiehdokaslistan.

Terminpoimintakokeen perusteella termiehdokkaan termiä vahvistavia ominaisuuksia ovat suuri esiintymien määrä tekstissä, vieraskielisyys ja lyhenteen olemassaolo. Termiehdokkaan termiä heikentäviä ominaisuuksia ovat variaatio ja yleiskielisyys. Manuaalisen ja puoliautomaattisen terminpoimintamenetelmän tärkein ero on, että ihminen tunnistaa termit ennen kaikkea semanttisten ominaisuuksien perusteella, käyttää hyödyksi tekstikontekstia, erikoisalan tietoa, kielellistä tietoa ja maailmantietoa sekä korjaa tekstin virheitä ja täydentää aukkoja. Merenkulun asiantuntijoiden, terminologisten menetelmien asiantuntijoiden ja opiskelijoiden kanssa toteutetun terminpoimintakokeen tulokset osoittivat, että erikoisalan tuntemuksen lisäksi terminpoiminnassa tarvitaan terminologian perusteiden tuntemusta. Puoliautomaattiset menetelmät eivät ainkaan lähitulevaisuudessa pysty termien semanttiseen tunnistukseen. Niiden tärkein etu manuaaliseen terminpoimintaan verrattuna onkin kyky suorittaa nopeasti ja virheettömästi ne tehtävät, jotka ne on ohjelmoitu suorittamaan.

Pelkkään terminpoimintaan tyytyminen ei kuitenkaan ole tarkoituksenmukaista, sillä termilistassa on vain osa tekstin terminologisesta tiedosta. Termilistan perusteella ei voi esimerkiksi varmistua siitä, mitkä termeistä tosiasiallisesti nimeävät samaa käsitettä. Tutkimukseni empiiriseen osuuteen kuuluvassa luvussa 7 olenkin tutkinut niin sanottuja terminologisia koettimia, joiden avulla teksteistä voidaan poimia termejä ja muuta terminologista tietoa. Tavoitteena oli tunnistaa suomalaisia ja venäläisiä koettimia sekä selvittää, soveltuvatko koettimet käsitetiedon puoliautomaattiseen poimintaan ja miten käsitetiedon poiminnan voi yhdistää terminpoimintaan. Koettimien tunnistuksen aloitin teksteistä, joista suomalaiset ja venäläiset asiantuntijat olivat tunnistaneet termit. Ensimmäisen korpustyökalun avulla konkordanssihaut tunnetuilla termeillä. Sitten tarkastelin konkordanssilistoista termien konteksteja selvittääkseni, olivatko kirjoittajat käyttäneet

termien yhteydessä ilmauksia, jotka ovat mahdollisia koettimia. Näin sain koottua alustavat koetinluettelot kummallakin kielellä. Tämän jälkeen tein manuaalisen koetinhaun laajemmalla suomen- ja venäjänkielisellä aineistolla ja täydensin alustavaa koetinluetteloa uusilla ehdokkailla. Lopuksi tein konkordanssihaut laajemmasta aineistosta koetinehdokkaista muotoiluilla hakuehdoilla. Konkordanssihakujen tulosten perusteella arvioin koetinehdokkaiden käyttökelpoisuutta asteikolla vahva, hyvä tai heikko sen mukaan, kuinka paljon ne tuottivat osumia ja kuinka suuri osa osumista oli valideja. En ole rajannut tutkimustani etukäteen tietynlaisiin koettimiin, esimerkiksi geneerisen suhteen tai verbikoettimiin, kuten joissakin aiemmissa tutkimuksissa on tehty, mutta en voi myöskään kieltää aiemman tutkimuksen ohjaavaa roolia. Englanninkielisillä aineistoilla tehtyjen tutkimusten tulokset ovat vaikuttaneet ainakin välimerkki- ja verbikoettimien valintaan. Olen asettanut tutkimukselleni kunnianhimoisen tavoitteen koota aineistostani kaikki erilaiset ja erilaista terminologisesti kiinnostavaa tietoa ilmaisevat koettimet. Aiemmat tutkimukset ovat olleet tuki, johon olen valinnassani nojannut. Valintani on kuitenkin subjektiivinen, ja toiset tutkijat olisivat todennäköisesti päätyneet samalla aineistolla erilaiseen ratkaisuun.

Konkordanssihaku osoitti, että noin puolet termien konteksteista sisältää jonkin koettimen. Kielestä riippumatta koettimet ovat usein preesensissä olevia verbejä ja sijaitsevat tekstissä tavallisesti joko välittömästi ennen termiä tai välittömästi sen jälkeen. Koettimet toimivatkin parhaiten yhdessä tunnettujen termien kanssa. Verbien lisäksi koettimina toimivat substantiivit, mutta myös eräät lyhenteet ja konjunktiot. Välimerkit toimivat koettimina vain yhdessä tunnetun termin kanssa, sillä konkordanssihaku ilman termiä tuottaa ylen määrin osumia, joista mahdollisesti vain pieni osa on valideja. Joskus koettimia on kaksi lähekkäin. Esimerkiksi käsitteen nimityksen koetin *käsité* saattaa esiintyä yhdessä jonkin toisen koettimen kanssa. Kahden koettimen käyttö saattaa vaikuttaa turhalta päällekkäisyydeltä, mutta käytön voi perustella sillä, että kirjoittaja on halunnut korostaa tiedon tärkeyttä ja varmistaa, että lukija ymmärtää, mitä hän on tarkoittanut.

Merenkulun turvallisuuden erikoisalalla suomalaiset ja venäläiset koettimet tuottavat tutkimukseni perusteella terminologista tietoa erityisesti keskenään kausaalisuhteessa olevista käsitteistä. Tämä selittyy sillä, että erikoisalana merenkulku kuuluu tekniikan ja sitä kautta luonnontieteiden alaan, jolle kausaalisuhteet ovat tunnusomaisia. Tapah-tuma on dynaaminen käsite. Siksi sen syyn ja seurauksen koettimina toimivat erityisesti verbit, suomessa verbit *aiheuttaa*, *aiheuttaa* ja *johtaa*, venäjässä verbit *привести*, *приводить* ja *вызвать*. Muita vahvoja syyn ja seurauksen koettimia ovat, kuten arvata saattaa, substantiivit *syy* ja *seuraus* sekä prepositio *из-за*. Myös toiminta on dynaaminen käsite. Sen tarkoituksen ja tuloksen vahvoina koettimina toimivat kuitenkin ennen kaikkea substantiivit *tavoite*, *tarkoitus*, *tehtävä* ja *результат*, mutta myös verbi *pyrkii* on vahva tarkoituksen koetin.

Tutkimallani merenkulun alalla koettimet ovat erittäin käyttökelpoisia myös kausaalisuhteisiin liittyvien toimintasuhteiden ja vaikuttavan tekijän poiminnassa. Vuorovaikutussuhteeseen liittyvän vaikuttavan tekijän vahvoja koettimia ovat verbit *vaikuttaa*, *parantaa* ja *vähentää*, substantiivi *влияние* sekä ilmaus *за счет*. Toimintasuhteisiin kuuluvia välinsuhteita voi hakea verbeillä *применять*, *использовать*, *служить* ja *предназначить* sekä substantiivilla *apu*, joka koettimina toimiessaan on muodossa *avulla*. Menetelmäsuhteen vahvoja koettimia ovat *käyttää* ja *metodi*, keinosuhteen koettimia ovat *keino*, *apu* ja *nytem* taivutusmuodossa *nytem*.

Koettimien avulla suomen- ja venäjänkielisistä merenkulun teksteistä voidaan poimia myös termejä ja muita käsitteen nimityksiä, geneerisiä suhteita ja koostumussuhteita sekä määritelmiä, joiden poimintaan koettimia onkin käytetty jo varsin laajasti eri

alojen englanninkielisillä aineistoilla. Käsitteen nimitysten hakuun parhaita koettimia ovat substantiivit *käsite*, *termi*, *nimi*, *nimitys*, *термин*, *понятие* ja *концепция* tai lyhenne *ns*. Suomenkielisistä vieruskäsitteen koettimista vahvin on konjunktio *ja* yhdessä yhdysmerkin kanssa. Määritelmiä tuottavat verbit *tarkoittaa* ja *представлять собой* sekä substantiivi *määritelmä*. Hierarkkisten geneeristen ja koostumussuhteiden koettimista vahvimpia ovat verbit *koostua*, *jakaa*, *относиться*, *отнести*, *состоять из* ja *включать в себя* sekä ilmaus *ja muut* eri taivutusmuodoissa. Sen sijaan synonymian paikallistamiseen ei tutkimukseni tuottanut yhtään vahvaa suomalaista tai venäläistä koetinta.

Tutkimukseni perusteella puoliautomaattisessa terminologisen tiedon poiminnassa koettimien avulla saavutetaan paras tulos, jos termit on poimittu ensin ja konkordanssihaku tehdään termien kanssa yhdessä, sillä termi on usein välittömästi ennen koetinta tai sen jälkeen. Suomen ja venäjän ”vapaa” sanajärjestys sallii kuitenkin ”ylimääräisten” sanojen sijoittumisen termin ja koettimen väliin, jolloin kaikki validit osumat eivät tulostu konkordanssilistaan, jos hakuehto on muotoiltu niin, että termin ja koettimen on oltava vierekkäin. Vahvat koettimet ovat kuitenkin käyttökelpoisia myös ilman tunnettua termiä, jolloin vapaa sanajärjestys ei vaikuta konkordanssihaun tulokseen. Koettimet toimivat hämmästyttävän samalla tavalla tutkimukseni suomenkielisessä ja venäjänkielisessä aineistossa. Tästä voi päätellä, että koettimet ovat erikoisalasidonnaisia. Toinen mahdollinen selitys on, että olen tiedostamatta etsinyt erikielisistä aineistoista samanlaisia koettimia.

Koetintutkimukseni tulokset osoittavat, että monet koettimiin liittyvät tekijät vaikeuttavat niiden käyttöä terminologisen tiedon puoliautomaattisessa poiminnassa. Eniten käsitetiedon tunnistamista vaikeuttaa ilmaisukeinojen runsaus. Yhtä käsitesuhdetta saatetaan ilmaista kymmenillä erilaisilla kielellisillä rakenteilla. Koettimien toiminnassa suomen ja venäjän kielellä näyttäisi olevan enemmän yhtäläisyyksiä kuin esimerkiksi suomen ja englannin kielellä. Suomen ja venäjän yhteisiä piirteitä ovat ainakin ”vapaa” sanajärjestys sekä nominien sijapäätteet ja verbien persoonapäätteet, jotka kaikki hankaloittavat konkordanssihaun hakuehdon muotoilua. Jos hakuehdossa päätteet korvataan katkaisumerkillä, suurin osa osumalistaan tulostuneista osumista on todennäköisesti roskaa. Jos taas sanat sisällytetään hakuehtoon päätteineen, lauseke muodostuu helposti monen kymmenen merkin mittaiseksi. Käsitesuhteen laadun selvittäminen vaatii koettimien ja käsitetiedon tarkastelua kontekstissaan, sillä sama koetin voi toimia erilaisten suhteiden ilmaisimena eikä koettimen perusteella välttämättä voi päätellä hierarkkisen suhteen suuntaa. Lisäksi tutkimukseni osoitti, että osa koettimista on kirjoittajakohdaisia ja mahdollisesti myös tekstilajikohdaisia. Koettimien käytöstä saatu hyöty riippuukin suoraan käytettävissä olevasta aineistosta. Parhaita terminologisen tiedon lähteitä ovat oppikirjat, tieteelliset julkaisut ja auktorisoitu aineisto. Tällaisesta aineistosta koettimet tuottavat parhaan tuloksen. Terminologiset koettimet ovat mitä ilmeisimmin myös alakohtaisia. Tutkimukseni osoitti, että merenkulun turvallisuuden alalla koettimet sopivat erityisesti kausaalisuhteiden hakuun. Tarkoitukseni onkin seuraavaksi tutkia kausaalisuhteiden koettimia uudella aineistolla.

Puutteista huolimatta olen tutkimukseni tulosten perusteella vakuuttunut, että koettimet ovat käyttökelpoisia terminologisen tiedon haussa. Niiden avulla teksteistä voidaan paikallistaa tietokasaumia, joiden perusteella lukija voi nopeasti päättää, onko teksti sopiva lähempään tarkasteluun. Uuteen alaan tutustuva kääntäjä tai tulkki voi koettimien avulla hakea käsitetietoa ja termien käyttöesimerkkejä nopeammin kuin lukiemalla teksti läpi. Koettimien toimintaperiaatteen tuntemisesta on hyötyä käsinpoiminnassa, mutta sitäkin enemmän koettimista on hyötyä automaattisessa tiedonhaussa, kun suuresta tekstimassasta voi nopeasti tulostaa terminologisesti kiinnostavat kontekstit.

Koettimet eivät kuitenkaan paljasta kaikkea terminologista tietoa tekstistä, koska aina niitä ei termien tai määritelmien yhteydessä esiinny. Tietotekniikan nopeasta kehityksestä huolimatta puoliautomaattiset menetelmät ovat sittenkin vain apukeinoja ja terminologisen tiedon poiminta vaatii yhä vielä runsaasti manuaalista työtä. Käsiteanalyysia ei pystytä tekemään tietokoneen avulla ainakaan lähitulevaisuudessa. Terminpoimintaan tarkoitetut puoliautomaattiset menetelmät eivät nekään vielä ole täysin kehittyneet. Tällä hetkellä markkinoilla olevat terminpoimintaohjelmat sopivat huonosti terminologisiin tarkoituksiin, sillä ne on kehitetty ensisijaisesti avainsanojen, ei teknisten termien, hakuun. Ohjelmien kielivalikoima on yleensä hyvin suppea ja hinta saattaa olla liian korkea esimerkiksi yksityisyrittäjänä työskentelevälle kääntäjälle. Terminologian näkökulmasta ohjelmien suurin puute on, että poimintatulos ei vastaa toiveita. Suurin osa termeistä jää pois poimintatuloksesta, koska ohjelmat poimivat pienifrekvenssiä termejä hyvin valikoivasti jos ollenkaan.

Tutkimukseni perusteella nykyisin tarjolla olevilla ohjelmilla saa parhaan tuloksen vaiheittaisella menetelmällä, jossa terminpoiminta aloitetaan ydintermeistä. Sopivilla asetuksilla saadaan tietystä tekstistä poimittua ydintermit millä tahansa terminpoimintaohjelmalla. Vaihtoehtoisesti ydintermien poiminnan voi tehdä konkordanssiohjelmalla käyttäen apuna koettimia ja sanafrekvenssejä. Luen ydintermeihin myös vieraskieliset termit, jotka on poimittava, koska niiden lähellä saattaa olla omakielinen termi, ja lyhenteet, koska avattu muoto saattaa olla lähellä. Termilistan perusteella on helppo päättää, käsitteleekö teksti aihetta, josta tietoa haetaan, ja jatketaanko poimintaa tällä tekstillä. Jos teksti osoittautuu hyväksi, poimintaa jatketaan konkordanssiohjelmalla, jossa hakuehdot muodostetaan termien ja koettimien yhdistelmistä (vrt. hakukone Googlen define-toiminto). Seuraavassa vaiheessa ydintermeillä sekä ydintermien ja koettimien yhdistelmillä haetaan synonyymeja, vastineita, ylä- ja alakäsitteiden nimityksiä, määritelmiä ja käsitesuhteita, joita tarvitaan käsitejärjestelmän rakentamiseen ja jotka tuottavat lisää termejä uutta hakua varten. Ydintermien lisäksi vaiheittaisen haun lähtötermeiksi sopivat alan yleiset tunnetut termit, joita ovat merenkulun turvallisuuden alalla esimerkiksi termit *meriliikenne* ja *laivaliikenne*, ja yleistieteelliset termit, kuten esimerkiksi *järjestelmä*. Näiden toistuvien vaiheiden kautta saadun terminologisen tiedon avulla pystytään mahdollisesti rakentamaan alustava käsitejärjestelmä, joka täydennetään muilla menetelmillä, esimerkiksi haastatteleamalla erikoisalan asiantuntijoita.

Terminologisen tiedon poimijoilla voi olla erilaisia tarpeita. Joissakin tapauksissa poimija haluaa koota tekstistä kaiken terminologisen tiedon. Esimerkiksi terminologille kaikki erikoisalan käsitteitä koskeva tieto on tärkeää, mutta myös erikoisalan opettaja tai tulkkaustehtävään valmistautuva tulkki haluaa muodostaa kokonaiskäsitteiden tietystä aiheesta ja haluaa siksi koota tekstistä kaiken aihetta koskevan tiedon. Heille ovat tärkeitä myös pienifrekvenssiset termit. Joissakin tapauksissa poimijalle riittää tiettyä käsitettä tai käsitejoukkoa koskeva täsmätieto. Esimerkiksi kääntäjä, opettaja tai erikoisalan asiantuntija saattaa hakea tekstistä ratkaisua juuri sillä hetkellä ajankohtaiseen käsite- tai termiongelmiaan. Kaikilla poimijoilla on kuitenkin yhteinen tavoite: saada kaikki haluttu asiaa koskeva tieto, mutta ei ylimääräistä, sillä hetkellä tarpeetonta tietoa.

Tutkimukseni perusteella automaattinen terminpoiminta tehostuisi, jos terminologisten koettimien käyttö yhdistettäisiin nykyisiin terminpoimintamenetelmiin. Mutta miksi tyytyä poimimaan teksteistä vain termejä, kun niissä on paljon muutakin terminologista tietoa, jonka poimintaa on mahdollista automatisoida. Mielestäni tutkimustulokset osoittavat selvästi, mihin suuntaan ohjelmia tulisi kehittää. Jätän tietoteknisten työkalujen kehittäjien ratkaistavaksi, voisiko terminpoiminnan ja konkordanssihaun yhdistää samaan työkaluun. Puoliautomaattisessa poiminnassa haun voisi koettimien avulla rajata vain tiettyyn tietoon tai kohdistaa kaikkeen tekstin terminologiseen tietoon. Tällainen



työkalu voisi olla tietokoneen työpöydällä valmiina käytettäväksi tarkasteltavan tekstin rinnalla. Ohjelman käyttäjä valitsisi ensin tekstin, josta hän haluaa poimia terminologista tietoa. Ensimmäisessä vaiheessa ohjelma tuottaisi tekstistä ydintermilistan, jonka perusteella käyttäjä päättäisi, sopiiko teksti tiedonhakuun. Seuraavaksi käyttäjä valitsisi termit, joiden nimeämistä käsitteistä hän haluaa lisätietoa. Lisäksi käyttäjä voisi valita, millainen valittuja käsitteitä koskeva tieto häntä kiinnostaa. Ohjelma käyttäisi tiedonhakuun konkordanssihakua, jossa hakuehdossa olisi termi ja koetin. Ohjelmaa voisi kehittää myös niin, että alakohtaiset haut olisivat mahdollisia. Tämä vaatisi erikoisalan tunnettujen ydintermien ja alakohtaisten koettimien tallentamista ohjelmaan. En osaa enustaa, miten tällainen ohjelma menestyisi markkinoilla, mutta jos ohjelma olisi helpokäyttöinen ja hankintahinnaltaan edullinen, se kiinnostaisi varmasti laajaa kieliammattilaisten joukkoa.

Tutkimuksessani tuli esiin, että terminpoiminnassa yksilölliset käsitykset termistä vaikuttavat termin valintaan enemmän kuin koulutus. Tulos ei vastannut ennakkoodotuksia. Ilmiötä olisikin kiinnostavaa tutkia lisää suurella joukolla koehenkilöitä, joilla on erilaiset koulutustaustat. Toinen terminpoimintaan liittyvä jatkotutkimusaihe on koettimien käytön yhdistäminen terminpoimintaan ja emotermimenetelmän kehittäminen. Terminologisiin koettimiin keskittyvässä tutkimuksessa jatkotutkimuksen aiheena kiinnostavimpia ovat verbikoettimet, koska tutkimukseni perusteella ne näyttävät tuottavat eniten terminologista tietoa (ks. jakso 7.6.1). Tutkimusaineistoni oli varsin suppea, eikä sen perusteella aina ollut mahdollista päätellä, mikä hakuehto tuottaa parhaan tuloksen. Haun tarkentamista konkordanssihaussa onkin tutkittava suuremmalla aineistolla. Koettimien yhteydessä ilmeni myös, että modaalisuus on yleistä. Tässä tutkimuksessa en ole paneutunut tarkemmin tähän ilmiöön, mutta sen yleisyyden vuoksi sitä kannattaisi tutkia myös terminologian näkökulmasta. Olisi mielenkiintoista selvittää, miten epävarmuutta ilmaistaan koettimien yhteydessä ja minkälaiseen terminologiseen tietoon modaalisuus liittyy.

Tutkimukseni antoi viitteitä siitä, että jotkin terminologiset koettimet saattavat olla alakohtaisia tai tekstilajikohtaisia. Alakohtaisuuden varmistamiseksi olisi koottava aineisto, jossa olisi monen eri kirjoittajan tekstejä vähintään kahdelta eri alalta. Tekstilajikohtaisuutta pitäisi tutkia aineistosta, jossa olisi eri tekstilajeihin kuuluvia eri kirjoittajien tekstejä. Tutkimusaineiston koostamisen vaikeus rajoittaa kuitenkin tutkittavien alojen valintaa. Helpoimmin aineiston koostaminen onnistunee aloilla, joilla eri tekstilajeihin kuuluvia tekstejä tuotetaan runsaasti ja monen kirjoittajan voimin.

Terminologian tulevaisuudesta voidaan olla kahta mieltä. Toisaalta terminologian tarve saattaa vähentyä, koska erikoisalat kansainvälistyvät, termistö yhtenäistyy ja englannin kielen valta-asema vahvistuu erikoisalojen viestinnän kielenä. Toisaalta voidaan väittää, että terminologia on tulevaisuuden ala, sillä kansalliset erikoiskielet eivät häviä minnekään, vaan juuri monikieliset ja monikulttuuriset yritykset tarvitsevat eniten termityötä. Esimerkiksi Nokia ja Kone ovat tällaisia kansainvälisiä yrityksiä, joissa tehdään laajamittaista termityötä. Myös standardoinnin ja sitä kautta sanastotyön tarve kasvaa maailmanlaajuisesti. Maailman taloustilanteesta riippumatta tiedon määrä kasvaa joka tapauksessa ja sen hallintaan on otettava kaikki keinot käyttöön. Tiedon tuottaminen ja arkistointi tapahtuu sähköisesti, eikä termien tai käsitetiedon hallinta vanhoilla manuaalisilla menetelmillä ole enää riittävän tehokasta. Terminologisen tiedon ja menetelmien merkitys korostuu, kun olennainen tieto on pystyttävä seulomaan epäolennaisesta. Yliopistojen tietopalvelut voisivat toimia tiennäyttäjinä ja ottaa hyötykäyttöön terminologisen tutkimuksen tulokset ja välineet. Yliopistollisessa koulutuksessa opiskelijoille olisi opetettava tiedonhaun ja -hallinnan menetelmiä monipuolisesti. Kielten osaamisala voisi toimia edelläkävijänä, mutta terminologisten menetelmien tuntemisesta



on hyötyä millä tahansa erikoisalalla. Terminologian perusteiden tulisi olla vähintään valinnainen kurssi ainakin kielten koulutusohjelmissa. Terminologian tutkimukselle on tarvetta, sillä tietotekniikan kehitys on nopeaa ja ohjelmien kehittäjät tarvitsevat tuekseen ajantasaista tutkimustietoa. Tutkimusten määrän kasvu antaisi terminologian alalle lisää painoarvoa tiedemaailmassa ja tekisi alaa tunnetuksi. Tutkimuksen ohessa voisi myös miettiä suomalaista alan nimitystä. Nythän terminologia yhdistyy suuren yleisön mielessä jonkin alan termistöön. Terminologiaa parempi nimitys olisi terminologiaoppi, termitiede tai termioppi, joka kertoisi lukijalle tai kuulijalle, että on kysymys tutkimus-  
alasta.

## LÄHDELUETTELO

### AINESLÄHTEET

- ato98 = *Alusturvallisuusohjelma 1998–2002*. Merenkululaitos. Saatavilla [www-muodossa](http://www.fma.fi/toiminnot/turvallisuus/alusturvallisuusohjelma.php)  
<URL:<http://www.fma.fi/toiminnot/turvallisuus/alusturvallisuusohjelma.php>>. 9.6.2003.
- Bezop = Skopkov, Konopel'ko & Vasil'eva 1994 = Снопков, В.И., Конопелько, Г.И. & Васильева, В.Б. 1994. *Безопасность мореплавания*. Москва: Транспорт.
- Eskos = Pričkin & Kljuev 2000 = Причкин, О. & Ключев, В. 2000. Единая система контроля над судоходством. *Морской флот* 5–6, 15–19.
- hv92 = Valkonen, H. 1992. Muuttuvat turvallisuussäädökset ja operatiiviset tarkastukset. Teoksessa J. Vainio (toim.) *Merenkulkualan VII teemapäivä. Turussa 9.3.1992 pidetyn teemapäivän esitelmät ja alustukset*. Turun yliopiston merenkulkualan koulutuskeskuksen julkaisuja B 40. Turku: Turun yliopisto, 11–15.
- hv98 = Valkonen, Heikki 1998. *Meriturvallisuus*. Esitelmä Turun yliopiston merenkulkualan koulutus- ja tutkimuskeskuksen Meriturvallisuusseminaarissa Helsingissä 5.2.1998.
- js94 = Sukselainen, J. 1994. Tekniikka ja merenkulun turvallisuus. Teoksessa J. Vainio (toim.) *Merenkulkualan IX teemapäivä. Helsingissä 7.3.1994 pidetyn teemapäivän esitelmät ja alustukset*. Turun yliopiston merenkulkualan koulutuskeskuksen julkaisuja B 65. Turku: Turun yliopisto, 3–12.
- kiv97 = *Komentosiltatyön inhimilliset virheet*. 1997. Helsinki: Merenkululaitos.
- kl98 = Larjo, K. 1998. Tekniikka ja merenkulkukulttuuri. Teoksessa J. Vainio (toim.) *VTS-päivien esitelmät 7.–8.12.1998*. Turun yliopiston merenkulkualan koulutus- ja tutkimuskeskuksen julkaisuja B 103. Turku: Turun yliopisto, 68–80.
- lat97 = Varsta, P. & Hanhirona, K. 1997. *Laivaturvallisuus*. M-217. Espoo: Teknillinen korkeakoulu.
- mv1 = AIS-järjestelmän Poseidon-kokeilut olivat menestys. 1998. *Meriväylä* 4. Saatavilla [www-muodossa](http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/498/index.php)  
<URL:<http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/498/index.php>>. 2.3.2005.
- mv2 = VTS laajenee vähitellen koko rannikolle. 2000. *Meriväylä* 3. Saatavilla [www-muodossa](http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/300/VTS_laajenee.html) <URL:  
[http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/300/VTS\\_laajenee.html](http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/300/VTS_laajenee.html)>. 9.3.2005.
- mv3 = Suomenlahden liikenteenvalvontajärjestelmän valmistelut pitkällä. 2001. *Meriväylä* 3. Saatavilla [www-muodossa](http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/301/vtmis.html)  
<URL:<http://www.fma.fi/palvelut/tietopalvelut/julkaisut/merivayla/301/vtmis.html>> 9.3.2005.
- mv4 = Kylmäluoma, R. 2003. Bothnia VTS vuoden kokemuksella: Alusliikennepalvelu auttaa merellä operoivia kohtaamaan oikeaan aikaan. *Meriväylä* 1, 25. Saatavilla [www-muodossa](http://www.fma.fi/media/merivayla/2003-1.pdf) <URL:<http://www.fma.fi/media/merivayla/2003-1.pdf>>. 9.3.2005.

- mv5 = Suomenlahden reittijakoalueet muuttuvat heinäkuussa. 2003. *Meriväylä* 2, 12–13. <http://www.fma.fi/media/merivayla/2003-2.pdf> (9.3.2005)
- mv6 = Montola, M. 2003. Uusi meriliikennekeskus otettiin käyttöön Helsingissä. *Meriväylä* 3, 20–21. Saatavilla [www-muodossa](http://www.fma.fi/media/merivayla/2003-3.pdf) <URL:<http://www.fma.fi/media/merivayla/2003-3.pdf>>. 9.3.2005.
- mv7 = Nyman, S.-H. 2003. VTS-palveluun liittyy monia lainsäädännöllisiä ja toiminnallisia haasteita. *Meriväylä* 3, 22–23. Saatavilla [www-muodossa](http://www.fma.fi/media/merivayla/2003-3.pdf) <URL:<http://www.fma.fi/media/merivayla/2003-3.pdf>>. 9.3.2005.
- mv8 = GOFREP vähentää onnettomuus- ja öljyvahinkoriskiä Suomenlahdella. 2004. *Meriväylä* 2, 8–9. Saatavilla [www-muodossa](http://www.fma.fi/media/merivayla/2004-2.pdf) <URL:<http://www.fma.fi/media/merivayla/2004-2.pdf>>. 9.3.2005.
- mv9 = Montola, M. & Hakkarainen, H. 2004. Suomenlahden liikenteen ilmoittautumisen alkoi. *Meriväylä* 3, 9–10. Saatavilla [www-muodossa](http://www.fma.fi/media/merivayla/2004-3.pdf) <URL:<http://www.fma.fi/media/merivayla/2004-3.pdf>>. 9.3.2005.
- mv10 = Heikonen, K. 2004. Alusten automaattinen tunnistusverkko kattaa pian kaikki Suomen vesialueet. *Meriväylä* 3, 11–12. Saatavilla [www-muodossa](http://www.fma.fi/media/merivayla/2004-3.pdf) <URL:<http://www.fma.fi/media/merivayla/2004-3.pdf>>. 9.3.2005.
- pw02 = Wihuri, P. 2002. *Meriliikenteen ohjausjärjestelmät VTS ja VTMISS*. Esitelmä Uudessakaupungissa 23.9.2002, 1–9.
- smr02 = Hänninen, S., Jalonen, R., Nyman, T., Palonen, A., Riska, K., Rytönen, J. & Sonninen, S. 2002. *Suomenlahden meriliikenteen riskitekijät*. Esiselvitys. Espoo: Teknillinen korkeakoulu.
- sms02 = Vainio, J. & Kajander, S. 2002. *Suomen merenkulun strategia. Ehdotus 2003–2012*. Liikenne- ja viestintäministeriön julkaisuja 46/2002. Liite 1. Helsinki: Liikenne- ja viestintäministeriö, 28–37.
- smt99 = *Suomen merikuljetusten turvallisuus. Tutkimusohjelma 1996–1998*. J. Sukse-lainen & R. Jalonen (toim.). Espoo: Valtion teknillinen tutkimuskeskus, 1999.

- Ahmad, K. 1994. Language Engineering and the Processing of Specialist Terminology. *The Language Engineering Convention/Journees du Genie Linguistique. Paris (France), 6–7 July 1994*. Organised by European Network in Language and Speech (ELSNET), EC2 (Nanterre, France), European Commission (DG XIII E), & French Ministries for Posts, Telecommunication and External Trade & Research. Saatavilla [www.muodossa](http://www.muodossa) <URL:<http://www.computing.surrey.ac.uk/ai/pointer/paris.html>>. 4.3.2004.
- Ahmad, K. & Fulford, H. 1992: Knowledge Processing 4: Semantic Relations and their Use in Elaborating Terminology. *Computing Sciences Report CS-92-7*. Guilford: University of Surrey.
- Ahmad, K. & Rogers, M. 2001: Corpus Linguistics and Terminology Extraction. Teoksessa S. E. Wright & G. Budin (eds.): *Handbook of Terminology Management. Vol. 2. Application-Oriented Terminology Management*. Amsterdam/Philadelphia: Benjamins, 725–760.
- Ahmad, K. & Salway, A. 1997: Safety and its Signification: A Case for a Language of Safety. *Fachsprache* 3–4(19), 94–110.
- Andreev 1999 = Андреев, Н. Д. 1999. Статистика-комбинаторные методы в теоретическом и прикладном языковедении. Teoksessa В. А. Татаринцов, *История отечественного терминоведения. Том 2. Направления и методы терминологических исследований. Очерк и хрестоматия. Книга 2*. Москва: Московский лицей, 122–138.
- Antia, B. E. 2000: *Terminology and Language Planning. An Alternative Framework of Practice and Discourse*. Amsterdam/Philadelphia: Benjamins.
- Antonova 1984 = Антонова, Т. И. 1984. Текстовые формулировки научного определения. *Вопросы языкознания* 1, 106–115.
- Arppe, A. 1995. *Term Extraction from Unrestricted Text*. Saatavilla [www.muodossa](http://www.muodossa) <URL:<http://www.lingsoft.fi/doc/nptool/term-extraction.html>>. 17.4.2002.
- Baker, M. 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7(2), 223–243.
- Barsalou, L. 1983. Ad hoc Categories. *Memory and Cognition* 11(3), 211–227.
- Bergenholtz, H. 1995. LGP and LSP. Teoksessa H. Bergenholtz & S. Tarp (eds.) *Manual of Specialized Lexicography. The Preparation of Specialized Dictionaries*. Amsterdam/Philadelphia: Benjamins, 16–20.
- de Bessé, B., Nkwenti-Azeh, B. & Sager, J. C. 1997. Glossary of Terms Used in Terminology. *Terminology* 4(1), 117–156.
- Bhatia, V. K. 2005. Genres in Business Contexts. Teoksessa A. Trosborg & P. E. F. Jørgensen (eds.) *Business Discourse*. Bern: Peter Lang, 17–39.
- Blair, D. 1990. *Language and Representation in Information Retrieval*. Amsterdam: Elsevier.
- Bondarenko 1992 = Бондаренко, И. В. 1992. *Английская терминология мореходства как предмет филологического исследования*. Диплома кандидата наук.

Автореферат диссертации на соискание ученой степени кандидата филологических наук. Москва: МГУ им. М. В. Ломоносова.

- Bourigault, D., Gonzales-Mullier, I. & Gros, C. 1996. LEXTER, a Natural Language Processing Tool for Terminology Extraction. Teoksessa Gellerstam, M. et al. (eds.). *EURALEX '96 Proceedings. I-II: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*. Göteborg: Göteborg University, 771–779.
- Bowker, L. 1996a. Learning from Cognitive Science: Developing a New Approach to Classification in Terminology. Teoksessa Gellerstam, M. et al. (eds.). *EURALEX '96 Proceedings. I-II: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*. Göteborg: Göteborg University, 781–787.
- Bowker, L. 1996b. Towards a Corpus-Based Approach to Terminography. *Terminology* 3(1), 27–52.
- Bowker, L. 1997. You Say "flatbed colour scanner", I Say "colour flatbed scanner". A Descriptive Study of the Influence of Multidimensionality on Term Formation and Use with Special Reference to the Subject Field of Optical Scanning Technology. *Terminology* 4(2), 275–302.
- Bowker, L. & Pearson, J. 2002. *Working with Specialized Language. A Practical Guide to Using Corpora*. London and New York: Routledge.
- Cabré Castells, M. T., Estopà Bagot, R. & Vivaldi Palatresi, J. 2001. Automatic Term Detection. A Review of Current Systems. Teoksessa D. Bourigault, C. Jacquemin & M.-C. L'Homme (eds.) *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: Benjamins, 53–87.
- Christensen, L. W. 2000. Danske verber som knowledge probes i terminologisk korpusarbejde. Teoksessa A. Nuopponen, B. Toft & J. Myking (red.). *I terminologins tjänst. Festskrift för Heribert Picht på 60-årsdagen*. Vaasa: Vaasan yliopiston julkaisu. Selvityksiä ja raportteja 59, 243–279.
- Citkina 1988 = Циткина, Ф. А. 1988. *Терминология и перевод. К основанию сопоставительного терминоведения*. Львов: Высшая школа.
- Condamines, A. & Rebeyrolle, J. 1998. CTKB: A Corpus-Based Approach to a Terminological Knowledge Base. Teoksessa D. Bourigault, C. Jacquemin & M.-C. L'Homme (eds.) *First Workshop on Computational Terminology, COMPUTERM '98, Montreal 15 August 1998*. Montreal: Université de Montréal, 29–35.
- Croft, W. & Cruse, D. A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Daille, B., Habert, B., Jacquemin, Ch. & Royauté, J. 1996. Empirical Observation of Term Variations and Principles for their Description. *Terminology* 3(2), 197–257.
- Danilenko 1971 = Даниленко, В. П. 1971. *Исследования по русской терминологии*. Москва: Наука.
- Dias, G., Guilloire, S. & Pereira Lopes, J. G. 2000. Benefiting from Multidomain Corpora to Extract Terminologically Relevant Multiword Lexical Units. Teoksessa U. Heid, S. Evert, E. Lehmann & C. Rohrer (eds.) *Proceedings of the Ninth Eura-*

*lex International Congress, EURALEX 2000, Stuttgart, Germany, August 8<sup>th</sup> – 12<sup>th</sup>, 2000. Vol. 1–2. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 339–350.*

- Dubuc, R. & Lauriston, A. 1997. Terms and Contexts. Teoksessa S. E. Wright & G. Budin (eds.) *Handbook of Terminology Management. Volume 1. Basic Aspects of Terminology Management*. Benjamins, Amsterdam/Philadelphia. 80–87.
- Felber, H. 1984. *Terminology Manual*. Paris: Unesco and Infoterm.
- Feliu, J. & Cabré, M. T. 2002. Conceptual Relations in Specialized Texts: New Typology and an Extraction System Proposal. Esitelmä konferenssissa *Terminology and Knowledge Engineering TKE '02*. Saatavilla [www-muodossa](http://www.muodossa) <URL:<http://tke2002.loria.fr/Doc/supports/feliu.ppt>>. 30.8.2004.
- Fontenelle, T. 2000. Extracting Phraseology for Content Analysis and Document Retrieval. Teoksessa U. Heid, S. Evert, E. Lehmann & C. Rohrer (eds.) *Proceedings of the Ninth Euralex International Congress, EURALEX 2000, Stuttgart, Germany, August 8<sup>th</sup> – 12<sup>th</sup>, 2000. Vol 1–2*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 351–358.
- Frantzi, K. T., Ananiadou, S. & Tsujii, J. 1998. The C-value/NC-value Method of Automatic Recognition for Multi-word Terms. Teoksessa C. Nikolaou & C. Stephanidis (eds.), *Lecture Notes in Computer Science: Natural Language Processing for Digital Libraries*. Heidelberg – New York – London: Springer-Verlag, 585–604.
- Fulford, H. 2001. Exploring Terms and their Linguistic Environment in Text. A Domain-Independent Approach to Automated Term Extraction. *Terminology* 7(2), 259–279.
- Gamper, J. & Stock, O. 1998/1999. Corpus-Based Terminology. *Terminology* 5(2), 147–159.
- Gaudin, F. 1993. *Socioterminologie: des problemes semantiques aux pratiques institutionnelles*. Mont-Saint-Aignan: Publications de l'Université de Rouen no 182.
- Gerd 1986 = Герд, А. С. 1986: *Основы научно-технической лексикографии*. Ленинград: Издательство Ленинградского университета.
- Golovin & Kobrin 1987 = Головин, Б. Н. & Кобрин, Р. Ю. 1987. *Лингвистические основы учения о терминах. Учебное пособие для студентов филологических специальных вузов*. Москва: Высшая школа.
- Gorbunov 2004 = Горбунов, Ю. И. 2004. *Тезаурусное моделирование французской грамматической терминологии*. Автореферат диссертации на соискание ученой степени доктора филологических наук. Санкт-Петербург: Российский государственный педагогический университет им. А. И. Герцена.
- Gorodetskaja 1993 = Городецкая, Л. А. 1993. Психолгнвистический эксперимент в терминоведении: понимание терминов специалистами и неспециалистами. Терминоведение и профессиональная лингводидактика. *Выпуск 1*, 30–38.
- Grinëv 2002 = Гринев, С. В. 2002. Актуальность проблемы когнитивного терминоведения. Научно-техническая терминология. *Выпуск 1*, 11–22.



- Grinëv 1993 = Гринев, С. В. 1993. *Введение в терминоведение*. Москва: Московский Лицей.
- Grinev, S. V. 2004. Terminology in the Era of Globalisation. Teoksessa Shelov, S.D., Leichik, V.M. (eds) *Russian Terminology Science (1992–2002)*. Vienna: Term-Net Publisher, 49–61.
- Grinsted, A. 2000. 'Knowledge probes' og eksempler – på jagt efter definitioner og begrebsrelationer i et korpus inden for området 'entrepreneurship'. Teoksessa A. Nuopponen, B. Toft & J. Myking (red.) *I terminologins tjänst. Festskrift för Heribert Picht på 60-årsdagen*. Vaasa: Vaasan yliopiston julkaisuja 59, 36–51.
- Guljakova 1984 = Гулякова, И. Г. 1984. *Судостроительная терминология современного русского языка*. Автореферат диссертации на соискание ученой степени кандидата филологических наук. Ленинград: Ленинградский орден на Ленина и ордена Трудового Красного Знамени государственный университет им. А.А. Жданова.
- Haarala, R. 1981. *Sanastotyön opas*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Hearst, M. A. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. *Actes de COLING-92*, Nantes, 1992.
- Heikkinen, V. 2003. *Tekstin kanssa. Kohti kielitieteellisen (kon)tekstintutkimuksen yleisöllä*. Saatavilla [www.muodossa.fi/tutkimus/tekstit/artikkelit/heikkinen\\_konteksti.pdf](http://www.muodossa.fi/tutkimus/tekstit/artikkelit/heikkinen_konteksti.pdf). 15.8.2003.
- Hirs, W. 1993. The Use of Terminological Principles and Methods in Medicine. Teoksessa H. B. Sonneveld & K. L. Loening (eds.) *Terminology. Applications in Interdisciplinary Communication*. Amsterdam/Philadelphia: Benjamins, 223–240.
- Hull, D. A. 1997. Automating the Construction of Bilingual Terminology Lexicons. *Terminology* 4(2), 225–244.
- ISO 12620:1999(E). *Computer Applications in Terminology – Data Categories*. Geneva: ISO.
- ISO 704:2000. *Terminology Work – Principles and Methods*. Second edition. Geneva: ISO.
- ISO 1087-1:2000(E/F). *Terminology Work – Vocabulary – Part 1: Theory and Application*. First edition. Geneva: ISO.
- ISO 1087-2:2000(E/F). *Terminology Work – Vocabulary – Part 2: Computer Applications*. First edition. Geneva: ISO.
- Ivina 2003 = Ивина, Л. В. 2003. *Лингво-когнитивные основы анализа отраслевых терминосистем*. Москва: Академический Проект.
- Jaakkola, S. 2002. *Kolmikielinen matkapuhelinsanasto – termipankkiohjelmien vertailua kääntäjän näkökulmasta*. Pro gradu -tutkielma. Turun yliopisto. Klassillisten ja romaanisten kielten laitos.
- Jacquemin, C. 1997. Recognition and Acquisition. Two Inter-Related Activities in Corpus-Based Term Extraction. *Terminology* 4(2), 245–273.
- Jacquemin, C. 2001. *Spotting and Discovering Terms through Natural Language Processing*. Cambridge (Massachusetts), London: MIT Press.

- Jokinen, A., Juhila K. & Suoninen E. 2004. *Diskurssianalyysin aakkoset*. 3. painos. Tampere: Vastapaino.
- Justeson, J. S. & Katz, S. M. 1995. Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering* 1(1), 9–27.
- Järvi, O. 2001. From Precise Terms to Fuzzy Words – From Bad to Worse in Terminology Science? *IITF Journal* 12(1–2), 85–88.
- Kageura, K. & Umino, B. 1996. Methods of Automatic Term Recognition: A Review. *Terminology* 3(2), 259–289.
- Kageura, K., Yoshioka, M., Koyama, T. & Nozue, T. 1998. Towards a Common Test-bed for Corpus-Based Computational Terminology. Teoksessa D. Bourigault, C. Jacquemin & M.-C. L’Homme (eds.) *First Workshop on Computational Terminology, COMPUTERM ’98, Montreal 15 August 1998*. Montreal: Université de Montréal, 81–85.
- Kageura, K., Yoshioka, M., Takeuchi, K., Koyama, T., Tsuji, K. & Yoshikane, F. 2000. Recent Advances in Automatic Term Recognition: Experiences from the NTCIR Workshop in Information Retrieval and Term Recognition. *Terminology* 6(2), 151–173.
- Kandelaki 1970 = Канделаки, Т. Л. 1970. *Значения терминов и системы значений научно-технических терминологий. Логические, лингвистические и историко-научные аспекты терминологии*. [Kustantajan tiedot puuttuvat] Москва.
- Karihalme, O. 1996. *Muotoilun teoriasanaston termistyminen*. Acta Wasaensia No 51. Kielitiede 10. Vaasa: Universitas Wasaensis.
- Katz, S. M. 1996. Distribution of Content Words and Phrases in Text and Language Modelling. *Natural Language Engineering* 2(1), 15–59.
- Kavanagh, J. 1995. *The Text Analyzer: A Tool for Extracting Knowledge From Text*. Master’s of Computer Science Thesis, University of Ottawa. Saatavilla [www.muodossa: <URL:http://www.csi.uottawa.ca/~kavanagh/Thesis>](http://www.csi.uottawa.ca/~kavanagh/Thesis). 10.1.2002.
- Kazarina 1998 = Казарина, С. Г. 1998. *Типологические характеристики отраслевых терминологий*. Краснодар: Издательство Кубанской государственной медицинской академии.
- Kemppanen, H. 2008. *Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Joensuun yliopiston humanistisia julkaisuja n:o 51. Joensuu: University of Joensuu.
- Kobrin 1989 = Кобрин, Р. Ю. 1989. *Лингвистическое описание терминологии как база концептуального моделирования в информационных системах*. Автореферат дис. д-ра филол. наук. Ленинград: ЛГУ.
- Koivisto-Alanko, P. 2000. *Abstract Words in Abstract Worlds. Directionality and Prototypical Structure in the Semantic Change in English Nouns of Cognition*. Helsinki: Yliopistopaino.
- Komarova 1991 = Комарова, З. И. 1991. *Семантическая структура специального слова и ее лексикографическое описание*. Свердловск: Издательство Уральского университета.

- Korpela, J. 2007. *Tyylilajin ja vaatimustason valinta. Millaisia tyylilajeja on?* Saatavilla [www-muodossa: <URL:http://www.cs.tut.fi/~jkorpela/kielenopas/3.3.html>](http://www.muodossa.fi/~jkorpela/kielenopas/3.3.html). Päivitetty 1.4.2007. 5.4.2007.
- Kostera, P. 1996. *Leksikologia saksa/suomi I. Wort für Wort. Oppi- ja harjoituskirja*. Helsinki: Finn Lectura.
- Kvitko, Lejčik & Kabancev 1986 = Квитко, И. С., Лейчик, В. М. & Кабанцев, Г. Г. 1986. *Терминоведческие проблемы редактирования*. Львов: Издательство при Львовском гос. университете.
- Kudashev 2007 = Кудашев, И. С. 2007. *Проектирование переводческих словарей специальной лексики*. Helsinki University Translation Studies Monographs 3. Helsinki: University of Helsinki.
- Laakso, M. 2003. *Automaattinen terminpoiminta: NaviTerm, LogiTerm 2.0 ja Trados 5 ExtraTerm –sovellusten arviointia*. Pro gradu -tutkielma. Turun yliopisto. Klassillisten ja romaanisten kielten laitos.
- Lahtinen, T. 2000. *Automatic Indexing: An Approach Using an Index Term Corpus and Combining Linguistic and Statistical Methods*. Department of General Linguistics Publications No. 34. Helsinki: University of Helsinki.
- Lakoff, G. 1987. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. Chicago and London: The University of Chicago Press.
- Laurén, C. & Picht, H. 1993. Vergleich der terminologischen Schulen. Teoksessa C. Laurén & H. Picht (toim.) *Ausgewählte Texte zur Terminologie*. Wien: TermNet, 493–539.
- Leitchik, V. M. & Shelov, S. D. 2004. Terminology: Where Is Russian Science Today? Teoksessa S. D. Shelov & V. M. Leitchik (eds.) *Russian Terminology Science (1992–2002)*. Vienna: TermNet Publisher, 15–48.
- Lejčik & Šelov 1990 = Лейчик, В. М. – Шелов, С. Д. 1990. *Лингвистические проблемы терминологии и научно-технический перевод. Часть II. Серия: Теория и практика научно-технического перевода. Обзорная информация. Выпуск 18*. Москва: Всесоюзный центр переводов.
- L’Homme, M.-C. 2000. Understanding Specialized Lexical Combinations. *Terminology* 6(1), 89–110.
- L’Homme, M.-C. 2003. Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivates. A Model for Specialized Lexicography. *International Journal of Lexicography* 16(4), 403–422.
- L’Homme, M.-C., Benali, L., Bertrand, C. & Lauduique, P. 1996. Definition of an Evaluation Grid for Term-Extraction Software. *Terminology* 3(2), 291–312.
- Lotte 1982 = Лотте, Д. С. 1982. *Вопросы заимствования и упорядочения иноязычных терминов и терминологических элементов*. Москва: Наука.
- Luk’janenkov & Sergeeva 1978 = Лукьяненко, К. Ф. & Сергеева, В. Н. 1978. *Частотный англо-русский словарь-минимум по судовождению*. Москва: Военное издательство министерства обороны СССР.
- Madsen, B. N., Pedersen, B. S. & Thomsen, H. E. 2001. *Defining Semantic Relations for OntoQuery*. Saatavilla [www-muodossa <URL:http://www.ontoquery.dk/publications/docs/Defining.doc>](http://www.ontoquery.dk/publications/docs/Defining.doc) 26.9.2008.

- Manerko 2000 = Манерко, Л. А. 2000. *Язык современной техники: ядро и периферия*. Рязань: РГПУ им. С. А. Есенина.
- Manerko 2003 = Манерко, Л. А. 2003. Истоки и основания когнитивно-коммуникативного терминоведения. Teoksessa В. Ф. Новодранова, Е. А. Никулина & Н. С. Новикова (ред.) *Лексикология Терминоведение Стилистика. Сборник научных трудов*. Москва–Рязань: Пресса, 120–126.
- Martin, W. 1992. On the Parsing of Definitions. Teoksessa H. Tammola, K. Varantola, T. Salmi-Tolonen & J. Schopp (eds.) *Euralex '92 Proceedings I-II. Papers Submitted to the 5<sup>th</sup> EURALEX International Congress on Lexicography in Tampere, Finland. Part I*. Tampere: University of Tampere, 247–256.
- Martin, W. J. R., Al, B. P. F. & van Sterkenberg, P. J. G. 1983. On the Processing of a Text Corpus. From Textual Data to Lexicographical Information. Teoksessa R. K. Hartmann (ed.) *Lexicography: Principles and Practice*. London: Academic Press, 77–87.
- Meyer, I. 2001. Extracting Knowledge-Rich Contexts for Terminography. A Conceptual and Methodological Framework. Teoksessa D. Bourigault, C. Jacquemin & M.-C. L'Homme (eds.) *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: Benjamins, 279–302.
- Mima, H. & Ananiadou, S. 2000. An Application and Evaluation of the C/NC-value Approach for the Automatic Term Recognition of Multi-word Units in Japanese. *Terminology* 6(2), 175–194.
- Montero-Martínez, S., García de Quesada, M. & Fuertes-Olivera, P. A. 2002. Terminological Phrasemes in OntoTerm®. A New Theoretical and Practical Approach. *Terminology* 8(2), 177–206.
- Muravitskaja & Perebejnos 1987 = Муравицкая, М. П. & Перебейнос, В. И. 1987. О лингвистическом обеспечении банка терминов. Teoksessa *Терминоведение и терминография в индоевропейских языках*. Сб. научных трудов. Владивосток: ДВО АН СССР, 83–91.
- Myking, J. 2001. Against Prescriptivism? The 'sociocritical' Challenge to Terminology. *IITF Journal* 12(1–2), 49–64.
- Niemelä, H. 2003. *De Motu. A Semantic and Pragmatic Framework of Motion in English and Finnish LST*. Helsinki University Translation Studies Monographs 1. Helsinki: University of Helsinki.
- Niemikorpi, A. 1996. *Liekepostista tuikeilmaisimeen ja sulhasesta kuraenkeliin. Erikoiskielten rakenteellisesta ja tyylillisestä vaihtelusta*. Vaasan yliopiston julkaisuja. Tutkimuksia 213. Kielitiede 34. Vaasa: Vaasan yliopisto.
- Nuopponen, A. 1993. Begriffssysteme und Textstruktur. Am Beispiel deutscher und finnischer Enzyklopädieartikel. Teoksessa H. Schröder (Hrsg.) *Fachtextpragmatik*. Forum für Fachsprachen-Forschung Band 19. Tübingen: Gunter Narr, 99–113.
- Nuopponen, A. 1994. *Begreppssystem för terminologisk analys*. Acta Wasaensia No 38. Språkvetenskap 5. Vasa: Universitas Wasaensis.
- Nuopponen, A. 2005. Concept Relations. An Update of a Concept Relation Classification. Teoksessa B. N. Madsen & H. E. Thomsen (eds.) *Terminology and Content Development. TKE 2005. 7<sup>th</sup> International Conference on Terminology and*

- Knowledge Engineering*. Copenhagen: Association for Terminology and Knowledge Transfer (GTW), 127–138.
- Nuopponen, A. 2008. Causal Concept Systems. Teoksessa B. N. Madsen & H. E. Thomsen (eds.) *Managing Ontologies and Lexical Resources. TKE 2008. 8<sup>th</sup> International Conference on Terminology and Knowledge Engineering*. Copenhagen: Institut for Internationale Sprogstudier og Vidensteknologi (ISV), 17–27.
- Nykänen, O. 1999. Sanastoprojektin vaiheet. Teoksessa K. Kuhmonen (toim.) *Toimikunnista termitalkoisiin. 25 vuotta sanastotyön asiantuntemusta*. Helsinki: Tekniikan Sanastokeskus, 62–71.
- Nykänen, O. & Kalliokuusi, V. 1999. Sanastotyön sanastoa. Teoksessa K. Kuhmonen (toim.) *Toimikunnista termitalkoisiin. 25 vuotta sanastotyön asiantuntemusta*. Helsinki: Tekniikan Sanastokeskus, 170–183.
- Nyström, L. 2000. *Alg, pytare och skridstång. Sälvfångstens och säljaktens terminologi i finlandssvenska folkmål*. Skrifter utgivna av Svenska litteratursällskapet i Finland nr 623. Helsingfors: Svenska litteratursällskapet i Finland.
- Opitz, K. 1979. Technical Dictionaries: Testing the Requirements of the Professional User. Teoksessa R. R. K. Hartmann (ed.) *Dictionaries and their Users: Papers from the 1978 (British association for applied linguistics) B.A.A.L. Seminar on Lexicography*. Exeter: University of Exeter, 89–95.
- Ozeki, S. 1987. Was ist der Begriff? Teoksessa H. Czap & C. Galinski (eds.) *TKE '87, Terminology and Knowledge Engineering. Proceedings*. Frankfurt/M: Indeks, 11–20.
- Parkkinen, J. 2002. *Hyvään verkkopalveluun! käytettävyysopas verkkoviestijöille*. Helsinki: Inforviestintä.
- Pasanen, P. 2003. Changing Terms in the Finnish VTS Terminology. *3<sup>rd</sup> International Conference on Maritime Terminology, ILTEC, Lisbon, 23<sup>rd</sup> and 24<sup>th</sup> June 2003*. Saatavilla CD-ROM-muodossa.
- Pasanen, P. 2005a. Terminologiset koettimet termien, käsitesuhteiden ja käsitepiirteiden ilmaisimina. Teoksessa K. Mård-Miettinen & N. Niemelä (toim.) *Erikoiskielet ja käännösteoria. VAKKI-symposiumi XXV, Vöyri 12.–13.2.2005*. Vaasan yliopiston käännösteorian, ammattikielten ja monikielisyyden tutkijaryhmän julkaisut N:o 32. Vaasa: Vaasan yliopisto, 290–298.
- Pasanen, P. 2005b. A Term List or a Noise List? How Helpful is Term Extraction Software when Finnish Terms are Concerned? Teoksessa B. N. Madsen & H. E. Thomsen (eds.) *Terminology and Content Development. TKE 2005. 7<sup>th</sup> International Conference on Terminology and Knowledge Engineering*. Copenhagen: Association for Terminology and Knowledge Transfer (GTW), 375–384.
- Pazienza, M. T. 1998/1999. A Domain-Specific Terminology-Extraction System. *Terminology* 5(2), 183–201.
- Pearson, J. 1996. The Expression of Definitions in Specialised Texts: A Corpus-Based Analysis. Teoksessa M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström, C. Rödger Pappmehl (eds.) *Euralex '96. Proceedings I-II Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg. Part II*. Göteborg: Göteborg University, 817–824.
- Pearson, J. 1998. *Terms in Context*. Amsterdam/Philadelphia: Benjamins.



- Pearson, J. 2000. Teaching Terminology Using Electronic Resources. Teoksessa S. P. Botley, A. M. McEnery & A. Wilson (eds.) *Multilingual Corpora in Teaching and Research*. Amsterdam – Atlanta: Rodopi, 92–105.
- Pekarskaja 1984 = Пекарская, Л. А. 1984. Разграничение терминов языка и терминов речи. Teoksessa *Термины в языке и речи*. Межвуз. сборник. Горький: ГГУ, 10–16.
- Picht, H. & Draskau, J. 1985. *Terminology: An Introduction*. Guildford: The University of Surrey.
- Picht, H. 2002. The Concept in Terminology. A Unit of Thought, Knowledge or Cognition? Esitelmä konferenssissa *Стандартизация терминологии, новые методы и результаты исследования терминологии разных областей знания*. Moskova 21.6.2002.
- Pilke, N. 2000. *Dynamiska fackbegrepp. Att strukturera vetande om handlingar och händelser inom teknik, medicin och juridik*. Acta Wasaensia No. 81. Språkvetenskap 15. Vaasa: Universitas Wasaensis.
- Pilke, N. 2002. The Concept and the Object in Terminology Science. *IITF Journal* 13(1-2), 7–26.
- Piotrovskij, Rahubo & Hažinskaja 1981 = Пиотровский, Р. Г., Рахубо, Н. П. & Хажинская, М. С. 1981. *Системное исследование лексики научного текста*. Кишинев: Штиинца.
- Pritchard, B. 2003. Multiword Lexical Units in Maritime Dictionaries. *3<sup>rd</sup> Conference on Maritime Terminology. ILTEC, Lisbon, 23<sup>rd</sup> and 24<sup>th</sup> June 2003*. Saatavilla CD-ROM-muodossa.
- Pälli, P. 2003. *Ihmisyhmä diskurssissa ja diskurssina*. Acta Universitas Tampereensis 910. Tampere: Tampereen yliopistopaino.
- Riggs, F. 1993. Social Science Terminology: Basic Problems and Proposed Solutions. Teoksessa H. B. Sonneveld & K. L. Loening (eds.) *Terminology. Applications in Interdisciplinary Communication*. Amsterdam/Philadelphia: Benjamins, 195–222.
- Rosch, E. 1978. Principles of Categorization. Teoksessa E. Rosch & B. Lloyd (eds.) *Cognition and Categorization*. NJ: LEA, Hillsdale, 27–48.
- Sager, J. 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Sager, J. 1994. *Language Engineering and Translation: Consequences of Automation*. Amsterdam/Philadelphia: John Benjamins.
- Salmi, R. 2003. Franchising ja tv-sovitin – näkökulmia termityöhön. *Terminfo* (2), 10–12.
- Sanastotyön käsikirja*. 1988. *Soveltavan terminologian periaatteet ja työmenetelmät. SFS-käsikirja 50*. Toimittanut Tekniikan Sanastokeskus. Helsinki: Suomen Standardisoimisliitto.
- Saukkonen, P. 1984. *Mistä tyyli syntyy. 2.1 Miten tyylilajit ovat syntyneet ja kehittyneet?* Saatavilla [www-muodossa](http://www.muodossa) <URL:<http://www.tritonia.fi/vanha/ov/wsoy/luku2.htm>>. 5.4.2007.



- Šelov 1998 = Шелов, С. Д. 1998. *Определение терминов и понятийная структура терминологии*. Санкт-Петербург: Издательство С.-Петербургского университета.
- Šelov 2001 = Шелов, С. Д. 2001. Терминоведение: семь вопросов и семь ответов по семантике термина. Teoksessa *Научно-техническая информация. Серия 2. Информационные процессы и системы* 2, 1–12.
- Sewangi, S. S. 2001. *Computer-Assisted Extraction of Terms in Specific Domains: The Case of Swahili*. Helsinki: University of Helsinki.
- Šilova 2005 = Шилова, Е. В. 2005. *Терминологическая дефиниция как метатекст в русскоязычной и англоязычной научно-технической литературе*. Автореферат диссертации на соискание ученой степени кандидата филологических наук. Екатеринбург: Уральский государственный педагогический университет.
- Sisseck, L. B. 2005. Terminological Knowledge Extraction – and Machine Learning for Danish. Teoksessa B. N. Madsen & H. E. Thomsen (eds.) *Terminology and Content Development. TKE 2005 7<sup>th</sup> International Conference on Terminology and Knowledge Engineering*. Copenhagen: Association for Terminology and Knowledge Transfer GTW, 385–395.
- Sierra, G. – McNaught, J. 2000. Design of an Onomasiological Search System: A Concept-Oriented Tool for Terminology. *Terminology* 6(1), 1–34.
- Soininen, P. 1998. *Terminhaun automatisointi*. Saatavilla [www.muodossa](http://www.muodossa.fi/URL:http://www.ling.helsinki.fi/~psoinine/lp.html) <URL:<http://www.ling.helsinki.fi/~psoinine/lp.html>>. 13.2.2002.
- Soininen, P. 1999a. *Sanastotyön tietotekniset apuvälineet*. Pro gradu -tutkielma. Helsingin yliopisto. Yleisen kielitieteen laitos.
- Soininen, P. 1999b. *Sanastotyön tietotekniset apuvälineet*. *Terminfo* (2), 7–10.
- SopS n:o 30/1977. Kansainväliset säännöt yhteentörmäämisen ehkäisemiseksi merellä, 1972*. Suomen asetuskokoelman sopimussarja, ulkovaltain kanssa tehdyt sopimukset 1977, N:o 30. Saatavilla [www.muodossa](http://www.muodossa.fi/URL:http://www.finlex.fi/fi/sopimukset/sopsteksti/1977/19770030/19770030_2) <URL:[http://www.finlex.fi/fi/sopimukset/sopsteksti/1977/19770030/19770030\\_2](http://www.finlex.fi/fi/sopimukset/sopsteksti/1977/19770030/19770030_2)>. 13.2.2002.
- Strehlow, R. 1997. Frames and the Display of Definitions. Teoksessa S. E. Wright & G. Budin (eds.) *Handbook of Terminology Management. Vol. 1. Basic Aspects of Terminology Management*. Amsterdam/Philadelphia: Benjamins, 75–79.
- Strehlow, R. 2001. The Role of Terminology in Retrieving Information. Teoksessa S. E. Wright & G. Budin (eds.) *Handbook of Terminology Management. Vol. 2. Application-Oriented Terminology Management*. Amsterdam/Philadelphia: Benjamins, 426–441.
- Suonuuti, H. 1999. Käsitemanalyysi työmenetelmänä. Teoksessa K. Kuhmonen (toim.) *Toimikunnista termitalkoisiin. 25 vuotta sanastotyön asiantuntemusta*. Helsinki: Tekniikan Sanastokeskus, 29–42.
- Swales, J. M. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Tariq, M., Manumaisupat, P., Al-Sayed, R. & Ahmad, K. 2003. Experiments in Ontology Extraction from Linguistic Resources. Teoksessa A. Todirascu & V Pallotta

- (eds.) *Proceedings of the International Workshop on Ontologies and Information Extraction, EUROLAN 2003*, 55–64.
- Tatarinov 1994 = Татаринов, В. А. 1994. *История отечественного терминоведения. Классики терминоведения: Очерк и хрестоматия*. Москва: Московский лицей.
- Tatarinov 1996 = Татаринов, В. А. 1996. *Теория терминоведения в трех томах. Том 1. Теория термина: История и современное состояние*. Москва: Московский лицей.
- Temmerman, R. 1996. Retrospect Lexicalisation: A Recurrent Phenomenon in the Lexicalisation Process of the Life Sciences. Teoksessa M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström, C. Rödger Pappmehl (eds.) *Euralex '96. Proceedings I–II Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg. Part II*. Göteborg: Göteborg University, 825–835.
- Temmerman, R. 2000. *Towards New Ways of Terminology Description. The Sociocognitive Approach*. Amsterdam/Philadelphia: Benjamins.
- Terminologian sanasto*. 2006. Helsinki: Sanastokeskus TSK.
- Uchimoto, K., Sekine, S., Murata, M., Ozaku, H. & Isahara, H. 2000. Term Recognition Using Corpora from Different Fields. *Terminology* 6(2), 233–256.
- Urizar, R., Ezeiza, N. & Alegria, I. 2000. Morphosyntactic Structure of Terms in Basque for Automatic Terminology Extraction. Teoksessa U. Heid ym. (eds.) *Proceedings of the Ninth Euralex International Congress, EURALEX 2000, Stuttgart, Germany, August 8<sup>th</sup> – 12<sup>th</sup>, 2000. Vol 1–2*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 373–381.
- Varantola, K. 1999. Korpukset sanastotyön lähteenä. Teoksessa K. Kuhmonen (toim.) *Toimikunnista termitalkoisiin. 25 vuotta sanastotyön asiantuntemusta*. Helsinki: Tekniikan Sanastokeskus, 130–138.
- Vehmas-Lehto, I. 2005. Terminologista tutkimusta opinnäytteinä. Teoksessa H. Yli-Jokipii (toim.) *Kielen matkassa multimediaan. Näkökulmia kääntämisen tutkimiseen ja opiskelemiseen*. Käännöstieteen laitoksen julkaisuja I. Helsinki: Helsingin yliopisto, 93–125.
- Vinokur 1939 = Винокур, Г. О. 1939. О некоторых явлениях словообразования в русской технической терминологии. Москва: *Труды Московского института истории, философии и литературы. Т. 5*, 3–54.
- Vivaldi, J. & Rodríguez, H. 2001. Improving Term Extraction by Combining Different Techniques. *Terminology* 7(1), 31–47.
- Wittgenstein, L. 1953. *Philosophical Investigations*. Transl. by G. Anscombe. Oxford: Basil Blackwell.
- Yli-Jokipii, H. 2006. Context, Genre and Translation. A View from Institutional Discourse. Teoksessa J. Tommola & Y. Gambier (toim.) *Translation and Interpreting – Training and Research. Traduction et interprétation – formation et recherche. Kääntäminen ja tulkkaus – koulutusta ja tutkimusta*. Turku: University of Turku, Department of English Translation Studies, 107–115.

## HAKEMISTO

- abstraktiotaso; 26, 57, 189  
adjektiivi; 61  
adjektiivimääräite; 78  
adverbi; 61  
agentti; 212  
agenttikausaalisuhde; 192  
aihealue; 155  
aineisto; 18, 148, 155  
aineiston laajuus; 20  
aito käyttöteksti; 19, 35, 37, 154  
ajatusyksikkö; 24  
alakäsite; 141  
algoritmi; 60, 120  
alkukielinen teksti; 19  
ammattikieli; 38  
ammattikielten tutkimus; 35  
ammattislangi; 38, 64  
antonymia; 51  
auktorisoitu aineisto; 152, 153  
automaattinen termintunnistus; 45  
avainsana; 45  
avainsanojen uuttaminen; 45  
deduktiivinen menetelmä; 68  
deskriptiivinen sanastotyö; 19, 36  
diakroninen termintutkimus; 64  
diakroninen termivariaatio; 64  
diakroninen tutkimus; 37  
dynaaminen konteksti; 65  
dynaaminen käsite; 145  
ekstralingvistinen konteksti; 65  
ellipsi; 63, 81  
elliptinen ilmaus; 30, 43, 86, 134  
emotermi; 48  
ensyklopedinen tieto; 142  
erikoisalakonteksti; 55, 65  
erikoisalakorpus; 68  
erikoisalatieto; 55, 71, 104  
erikoisalatunniste; 48  
erikoiskieli; 17, 32, 38  
frekvenssi; 46, 48, 56, 112, 116, 118, 121, 123, 124, 125, 126, 131  
funktiosuhde; 143, 144, 192  
geneerinen käsitejärjestelmä; 15  
geneerinen käsitesuhde; 142, 144, 176, 225  
genetiivimääräite; 83  
hakuehto; 159, 262  
hakulauseke; 148  
harmonisointi; 40  
heikko koetin; 160  
heuristiikka; 36  
homonymia; 30  
hyvä koetin; 160  
ihannetermi; 29  
indeksitermi; 28, 124  
indeksointi; 45, 106  
induktiivinen menetelmä; 68  
informaatiotiede; 36  
internationalismi; 28  
intuitio; 44  
joukkomääritelmä; 141  
kansainvälinen laina; 40  
kansainvälinen tieteellinen termi; 28  
kategorisaatio; 32, 36  
kausaalisuhde; 192  
keinokieli; 39  
keinosuhde; 209, 211, 258  
kielellinen konventio; 41  
kielellinen tieto; 71  
kielenhuolto; 34  
kielenohjailu; 25  
kirjallinen lähde; 151  
klassinen luokittelu; 24  
knowledge probe; 51  
koehenkilö; 69  
koetintutkimus; 146  
kognitiivinen taso; 36  
kognitiivinen terminologia; 31, 38  
kognitiotiede; 36  
kognitioyksikkö; 24  
kommunikatiivinen taso; 36  
kommunikatiivinen terminologia; 36  
konkordanssi; 159  
konkordanssitoiminto; 159  
kontaktisuhde; 143  
konteksti; 24, 37, 65  
kontekstimääritelmä; 141, 171  
kontekstisidonnaisuus; 30  
koostumussuhde; 143, 144, 183, 234  
koostumussuhteinen käsitejärjestelmä; 15, 143  
kopula; 264  
korpuslingvistiikka; 49  
korpustyökalu; 20, 159  
kvasisynonyymi; 28  
käsite; 22, 23

käsiteanalyysi; 15, 34, 50  
 käsitehierarkia; 26  
 käsitekehys; 139  
 käsiteluokka; 24  
 käsitepiirre; 145, 193  
 käsitesuhde; 141, 142  
 käsitetiedon poiminta; 16  
 käsitetieto; 14, 22, 43, 141, 142  
 käsitteen ala; 143  
 käsitteen elinkaari; 25, 27, 267  
 käsitteen nimitys; 161  
 käsitteen sisältö; 143  
 käytettävyys; 73  
 käännösteksti; 19  
 lausetermi; 58  
 lingvistinen konteksti; 65  
 lingvistinen taso; 36  
 lingvistinen terminpoimintamenetelmä;  
 47, 111, 139  
 looginen käsitesuhde; 143  
 luonnollinen kieli; 39  
 lyhennekoetin; 269  
 lyhennetermi; 59, 62, 78, 81, 84, 86, 98,  
 134, 135  
 maailmantieto; 65, 71  
 manuaalinen koetinhaku; 159  
 manuaalinen terminpoiminta; 43, 45,  
 128  
 menetelmäsuhde; 210, 254  
 merenkulun erikoisala; 17  
 merenkulun kieli; 40  
 merenkulun suomi; 42  
 modaaliapuverbi; 165, 248, 267  
 monisanainen termi; 46, 57  
 morfologinen variantti; 29  
 morfosyntaktinen termivariaatio; 63, 85  
 Moskovan koulukunta; 33  
 motivoituneisuus; 30  
 määrite; 59, 77  
 määritelmä; 22, 55, 141  
 määritelmätieto; 141  
 negatiivinen koetin; 148  
 nollafrekvenssi; 113  
 nollafrekvenssinen termi; 90  
 normatiivinen sanastotyö; 34, 36  
 objektisuhde; 213  
 olennainen käsitepiirre; 141  
 olio; 23  
 oliokäsite; 145  
 ominaiskäsite; 23  
 ominaisuus; 23  
 ominaisuuskäsite; 145  
 ontologia; 38  
 ontologinen käsitesuhde; 143  
 ontologinen luokka; 23  
 operaatiokieli; 39  
 ortografinen termivariaatio; 63  
 osittaissyonymia; 62  
 perinteinen terminologia; 33, 35  
 perustason käsite; 26  
 pienifrekvenssinen termi; 90, 97  
 piilotermi; 50, 87, 98, 115, 122, 125  
 polysemia; 30  
 Prahan koulukunta; 33  
 produktiivisuus; 160  
 prototyypiteoria; 24, 26  
 psykolingvistiikka; 36  
 puoliautomaattinen terminpoiminta; 44,  
 129  
 purskeisuus; 57  
 pysyvyys; 57, 73  
 päätermi; 28  
 riittävä piirre; 24  
 saanti; 73, 111, 121, 128, 137, 160  
 sanakirja; 18, 152  
 sanaliitto; 79, 84  
 sanaliittotermi; 57, 98  
 sanastoaineisto; 34  
 sanastostandardi; 30  
 sanastotyö; 13, 19  
 sanastotyömenetelmä; 14  
 sanastotyön lähde; 152  
 sekakoosteinen käsitejärjestelmä; 189  
 semanttinen kolmio; 22  
 semanttinen termivariaatio; 63, 81, 85,  
 134, 136  
 seuraus; 192, 244  
 sisäinen käsitepiirre; 145  
 sisältömääritelmä; 141, 171  
 sosiaalinen ulottuvuus; 35, 40  
 sosiologinen luokittelu; 25  
 sosioterminologia; 35  
 staattinen konteksti; 65  
 standardi; 152  
 standardisointi; 26, 33, 34, 54  
 substantiivi; 60  
 substantiivikoetin; 268  
 substantiivilauseke; 47, 60  
 substantiivimäärite; 77  
 substantiivitermi; 98

sulkulista; 47, 48, 72, 107, 140  
suullinen lähde; 151, 152  
suure; 23  
suurifrekvenssinen termi; 88, 97  
synkronian periaate; 37  
synkroninen termivariaatio; 64  
synkroninen tutkimus; 156  
synonymia; 26, 62, 133, 166, 219  
synonyymi; 28, 30, 62  
syntaktinen termivariaatio; 63, 80, 85, 135  
syy; 192, 244  
sähköinen lähde; 151  
sääösteksti; 19  
tapahtuma; 23  
tapahtumakausaalisuhte; 192  
tarkkuus; 73, 111, 121, 128, 137, 160  
tarkoite; 22  
tarkoitus; 192, 193, 244  
tekninen termi; 29, 124  
teksti; 19, 151  
tekstiaineisto; 19  
tekstiarkisto; 149  
tekstikonteksti; 66  
tekstikorpus; 45, 149  
tekstilaji; 151  
tekstin funktio; 155  
tekstin julkaisumuoto; 151  
temporaalinen käsitejärjestelmä; 143  
temporaalinen käsitesuhde; 143  
termi; 22, 27  
termiehdokas; 43, 70  
termiehdokaslista; 20, 50, 70  
termijärjestelmä; 29  
termikaava; 44, 47, 50, 61, 78, 84, 113, 116, 118, 123, 124, 125, 126, 132  
termikontekstisana; 49  
termilista; 20, 43, 68, 75  
termin lingvistinen piirre; 54  
termin lyhyt muoto; 28  
termin pitkä muoto; 28  
termin pituus; 112, 121, 124  
terminologi; 34, 45, 139  
terminologia; 32  
terminologinen koetin; 16, 52, 146  
terminologinen määritelmä; 141  
terminologinen tutkimus; 17  
terminologinen yksikkö; 30  
terminpoiminta; 37, 43  
terminpoimintamenetelmä; 14, 16  
terminpoimintaohjelma; 46  
termintunnistusohjelma; 106  
termistö; 32  
termivariaatio; 26, 62, 80, 136  
termivariantti; 30, 63, 134  
termiys; 28, 46, 54, 73  
termiyspisteet; 73, 107, 114, 120, 122, 124, 125, 127, 137  
tetraedrimalli; 22  
tiedeyhteisön hyväksymä aineisto; 152, 153  
tiedonhaku; 45  
tietokonelingvistiikka; 36  
tietämysyksikkö; 24  
tilastollinen terminpoimintamenetelmä; 46, 139  
toiminta; 23  
toimintasuhde; 192, 209  
tulos; 192  
turvallisuuskieli; 17  
tyyli; 156  
täsmäkorpus; 149  
UDK-luokitus; 155  
ulkoinen käsitepiirre; 145  
vahva koetin; 160  
vaikuttava tekijä; 204, 250  
vaikutussuhde; 143, 144, 192  
vajaa määritelmä; 142, 171  
vakiintumaton aineisto; 152, 153  
valetermi; 50, 87, 93, 101, 117, 123, 126  
variaatio; 133  
vastine; 81, 167  
verbi; 60  
verbikoetin; 264  
verbitermi; 80  
Wienin koulukunta; 33  
vieruskäsite; 189  
vieruskäsitesuhde; 144, 241  
viestintätilanne; 25, 36, 40, 154  
vuorovaikutussuhde; 192, 204  
välinesuhde; 192, 209, 211, 255  
välttämätön piirre; 24  
ydintermi; 89, 98, 137  
yhdistelmämenetelmä; 48  
yhdysmerkki; 77, 116  
yhdyssana; 79  
yhdyssanatermi; 58  
yhdyssubstantiivi; 79  
yksilökäsite; 23, 103

yksimerkityksisyys; 30  
yksisanainen termi; 46, 48  
yleiskieli; 39  
yleiskoetin; 147, 263  
yleiskäsite; 23, 28

yleisnimi; 28  
yläkäsite; 141  
ylätason käsite; 26  
ymmärryksikkö; 24



## LIITTEET

### LIITE 1

#### Tutkimusaineiston kuvaus

Lyhenne	Kieli	Julkaisu- muoto	Tekstilaji	Viestintätilanne	Funktio	Julkai- su- vuosi	Sana- luku	Aihealue
ato98	fi	sähköinen	periaateohjelma	asiantuntijalta asiantuntijalle	informatiivinen	1998	700	alusturvallisuus
hv92	fi	painettu	esitelmä	asiantuntijalta asiantuntijalle	informatiivinen	1992	647	merenkulun turvallisuus
hv98	fi	julkaise- maton	esitelmä	asiantuntijalta asiantuntijalle	informatiivinen	1998	1 076	merenkulun turvallisuus
js94	fi	painettu	esitelmä	asiantuntijalta asiantuntijalle	informatiivinen	1994	2 085	merenkulun turvallisuus
kiv97	fi	painettu	tutkimusraportti	asiantuntijalta asiantuntijalle	informatiivinen	1997	18 504	komentosiltatyö
kl98	fi	painettu	esitelmä	asiantuntijalta asiantuntijalle	informatiivinen	1998	3 893	merenkulku- kulttuuri
lat97	fi	painettu	tutkimusraportti	asiantuntijalta asiantuntijalle	informatiivinen	1997	709	alusturvallisuus
mv1– mv10	fi	painettu / sähköinen	artikkeli	puoliasian- tuntijalta puoliasian- tuntijalle	informatiivinen	1998– 2004	5 329	merenkulun turvallisuus
pw02	fi	julkaise- maton	esitelmä	asiantuntijalta asiantuntijalle	informatiivinen	2002	2 366	Meriliikenteen ohjausjärjestelmät
smr02	fi	painettu	esiselvitys	asiantuntijalta asiantuntijalle	informatiivinen	2002	23 349	meriliikenteen riskit
sms02	fi	painettu	strategiaehdotus	asiantuntijalta asiantuntijalle	ohjaava	2002	2 881	merenkulku
smt99	fi	painettu	tutkimusraportti	asiantuntijalta asiantuntijalle	informatiivinen	1999	11 497	merikuljetusten turvallisuus
Bezop	ru	painettu	oppikirja	asiantuntijalta puoliasian- tuntijalle	informatiivinen	1994	38 232	merenkulun turvallisuus
Eskus	ru	painettu	artikkeli	asiantuntijalta asiantuntijalle	informatiivinen	2000	2 794	Meriliikenteen ohjausjärjestelmät

## LIITE 2

Ohjekirje suomenkielisille koehenkilöille

Päivi Pasanen  
Helsingin yliopiston käännöstieteen laitos  
PL 94  
45101 KOUVOLA  
Puh. (05) 8252 220  
Sähköposti: paivi.pasanen@helsinki.fi

Syyskuu 2003

Hyvä vastaanottaja,

Ohessa on meriliikenteen ohjausjärjestelmiä käsittelevä teksti. Pyydän sinua merkitsemään tekstiin, mitkä sanoista mielestäsi ovat merenkulun termejä. Alleviivaa termit silloinkin, kun ne toistuvat.

### Ohjeita:

Termi on jollakin erikoisalalla (tässä tapauksessa siis merenkulun erikoisalalla) yleiskäsitteen nimityksenä käytettävä kielellinen ilmaus. Se on alalla yleisesti käytössä ja tunnettu.

Termi voi olla

- **yksi sana** (esim. suuntima)
- **yhdyssana** (esim. konealus)
- **kahden tai useamman sanan sanaliitto**  
(esim. ohjailukyvytön alus, ympäri näköpiirin näkyvä valkoinen valo)
- **lyhenne** (esim. ISM).

Myös verbi voi olla termi tai termin osa.

Esimerkki: hinata; sivuuttaa; pysäyttää koneisto

Saman käsitteen nimityksinä voi olla useita termejä.

Esimerkki: ympäristönseuranta TAI ympäristön tilan seuranta  
ISM TAI turvallisuusjohtamisjärjestelmä

Voit käyttää yhtenä valintaperusteena sitä, mitkä sanat, sanaliitot tai lyhenteet mielestäsi pitäisi olla merenkulun sanakirjassa.

Alleviivaa termi yhtenäisellä alleviivauksella. Jos termiin kuuluu mielestäsi kaksi tai enemmän sanoja, älä katkaise alleviivausta sanavälin kohdalla. Jos sanat ovat yksinäänkin termejä, käytä kaksoisalleviivausta.

**Esimerkki:** merenkulun turvallisuus

Jos termi jatkuu seuraavalle riville, piirrä rivin loppuun yhdistävä kaari.

**Esimerkki:** ohjailukyvytön  
alus

Jos termissä on enemmän kuin yksi sana ja sanojen välissä on mielestäsi termiin kuulumaton sana tai sanoja, alleviivaa termiin kuuluvat sanat ja yhdistä viivat katkoviivalla.

**Esimerkki:** näkyvyyden tarkempi arviointi  
\_\_\_\_\_

Jos kahdella termillä on yhteinen osa, alleviivaa se kaksoisalleviivauksella ja yhdistä erillään olevat yhteen kuuluvat osat katkoviivalla toisiinsa. Tarvittaessa alleviivauksia voi olla kolmekin allekkain.

**Esimerkki:** aluksen varustelutaso ja kunto  
\_\_\_\_\_

**Esimerkki:** asema- tai merkkivalo  
\_\_\_\_\_

**Esimerkki:** öljytankkerin ja matkustaja-aluksen ajolinjat  
\_\_\_\_\_  
\_\_\_\_\_

Paperin reunoille (tai eri paperille) voit kirjoittaa huomautuksia, kommentteja ja kysymyksiä.

Vaivannäöstä kiittäen

Päivi Pasanen

Ohjekirje venäjänkielisille koehenkilöille

Пяйви Пасанен  
Хельсинкский университет  
Отделение переводоведения  
П/Я 94  
ФИН 45101 КОУВОЛА  
Тел. +358 5 8252 220  
e-mail: paivi.pasanen@helsinki.fi

Август 2003

**Уважаемый получатель!**

В приложении - текст *Единая система контроля над судоходством* (22 стр.). Пожалуйста, прочитайте текст и подчеркните все слова, которые, по Вашему мнению, являются терминами. Подчеркните все термины, даже повторяющиеся, спонтанно. На первой странице укажите, пожалуйста, изучаете ли вы иностранные языки или судовождение.

**Дополнительные инструкции:**

Термин – это языковое выражение, соответствующее определенному понятию языка для специальных целей (в данном случае языка мореходства). Термины широко распространены и общеизвестны в данной области.

Термином может быть:

- **слово** (например, пеленг)
- **словосочетание** (парусное судно, судно с механическим двигателем)
- **сокращение** (ТЭО).

Можете подчеркнуть также глаголы и словосочетания с ними, которые, по Вашему мнению, имеют терминологическое значение.

**Примеры:** маневрировать, застопорить движители

Одному понятию может соответствовать несколько терминов.

**Примеры:** фактор среды, фактор окружающей среды  
ТЭО, технико-экономическое обоснование

В качестве критерия терминологичности можно также использовать следующий: терминами, скорее всего, являются слова и словосочетания, которые целесообразно поместить в морской словарь.

Подчеркните термины сплошной линией. Если, по Вашему мнению, термин состоит из двух или более слов, не прерывайте линию между словами. Если, по Вашему мнению, слова также самостоятельно являются терминами, подчеркните их двойной линией и оставьте пробел между отдельными терминами.

**Пример:** безопасность мореплавания

Если часть термина перенесена на следующую строку, поставьте в конце строки соединительную дугу.

**Пример:** изменение скорости

Если термин состоит из двух или более слов и между ними имеются слова, которые, по Вашему мнению, не являются частью термина, подчеркните слова, которые являются частями термина, и соедините их прерывистой линией.

**Пример:** затруднять движение такого судна

Если Вы считаете, что какое-либо слово относится одновременно к двум разным терминам, подчеркните общую часть двойной линией и соедините разделенные части термина прерывистой линией. При необходимости, подчеркните тремя линиями.

**Пример:** Изменение курса, а также скорости

**Пример:** узкий проход или фарватер

Замечания, комментарии и вопросы можно писать на полях или на отдельном листке.

Благодарю Вас за помощь,

Пяйви Пасанен

## LIITE 4

Suomalaisten asiantuntijoiden valitsemat termit tekstistä *pw02* (189 termiä)

termi	f	sanaluku	kaava
ahtaus	1	1	N
Allied Services	2	2	en
alueellinen liikenteen informaatiopalvelu	2	3	A N N <sub>c</sub>
aluemi	1	1	N <sub>c</sub>
aluksen päällikkö	1	2	N N
alus	5	1	N
alusraportointi	1	1	N <sub>c</sub>
alusten turvallisuus	1	2	N N
asema	2	1	N
ECDIS	1	1	Abb
ECDIS korjaus	1	2	Abb N
EEZ/LRT-järjestelmä	1	1	AbbN
elektroninen tietojenvaihto	2	2	A N <sub>c</sub>
etsintä	2	1	N
etsintä ja pelastus	0	3	N C N
etsintäviranomais	0	1	N <sub>c</sub>
Euroopan ympäristörannikkovartiosto	1	2	N N <sub>c</sub>
European Environmental Coastguard	1	3	en
Geographical Information System	1	3	en
HAZMAT-direktiivi	1	1	AbbN
hinaus	0	1	N
hinauspalvelu	1	1	N <sub>c</sub>
horisontaalinen integraatio	1	2	A N
horisontaalinen tietojenvaihto	1	2	A N <sub>c</sub>
huono näkyvyys	1	2	A N
huono näkyvyysolosuhde	2	2	A N <sub>c</sub>
IALA	6	1	Abb
IALA VTS Manual	1	3	en
IMCO	1	1	Abb
IMO	7	1	Abb
informaatiopalvelu	6	1	N <sub>c</sub>
integroitu tutkaverkko	1	2	A N <sub>c</sub>
integroitu verkko	1	2	A N
jakelu	1	1	N
kansainvälinen merenkulkujärjestö	2	2	A N <sub>c</sub>
karttakorjauspalvelu	1	1	N <sub>c</sub>
kokonaiskuljetusketju	1	1	N <sub>c</sub>
kuljetus	2	1	N
kuljetusketju	0	1	N <sub>c</sub>
kulku	2	1	N
käyntisatama	1	1	N <sub>c</sub>
laitejärjestelmä	2	1	N <sub>c</sub>
laiva	14	1	N
laivakäyttöinen tutka	1	2	A N
laivaliikenne	6	1	N <sub>c</sub>
laivaliikenteen ohjaus	2	2	N <sub>c</sub> N
laivaliikenteen ohjaus- ja informaatiojärjestelmä	1	4	N <sub>c</sub> N C N <sub>c</sub>
laivaliikenteen ohjaus- ja informaatiopalvelu	2	4	N <sub>c</sub> N C N <sub>c</sub>
laivaliikenteen ohjauspalvelu	1	2	N <sub>c</sub> N <sub>c</sub>



laivan käsittely	0	2	N N
laivan sijainti	1	2	N N
laivojen ilmaiseminen	1	2	N N
laivojen jäljittäminen	0	2	N N
laivojen tunnistaminen	0	2	N N
lastaus	0	1	N
lastaussuunnitelma	1	1	N <sub>c</sub>
lasti	3	1	N
lastin käsittely	1	2	N N
liikenne	9	1	N
liikennetiedosto	1	1	N <sub>c</sub>
liikenneturvallisuus	1	1	N <sub>c</sub>
liikenteen informaatiopalvelu	2	2	N N <sub>c</sub>
liikenteen järjestäminen	1	2	N N
liikenteen ohjaus	2	2	N N
liikenteenohjausjärjestelmä	2	1	N <sub>c</sub>
logistinen palvelu	1	2	A N
Long Range Tracking	1	3	en
luotsi	3	1	N
luotsipalvelu	0	1	N <sub>c</sub>
lähivesi	1	1	N <sub>c</sub>
maalle rakennettava tutka-asema	1	3	N A N <sub>c</sub>
maalle rakennettu tutka-asema	2	3	N A N <sub>c</sub>
Maritime Industries Forum	1	3	en
merenkulkija	2	1	N <sub>c</sub>
merenkulku	1	1	N <sub>c</sub>
merenkulkuala	2	1	N <sub>c</sub>
merenkulkukoulutus	1	1	N <sub>c</sub>
merenkulkupalvelu	1	1	N <sub>c</sub>
Merenkulun foorumi	1	2	N <sub>c</sub> N
merikuljetus	7	1	N <sub>c</sub>
meriliikenne	11	1	N <sub>c</sub>
meriliikenteen ohjaus- ja informaatiopalvelu	1	4	N <sub>c</sub> N C N <sub>c</sub>
meriliikenteen ohjausjärjestelmä	4	2	N <sub>c</sub> N <sub>c</sub>
meriliikenteen turvallisuus	2	2	N <sub>c</sub> N
meriliikenteenohjausjärjestelmä	1	1	N <sub>c</sub>
meritse tapahtuva kuljetus	1	3	Adv A N
meriympäristö	2	1	N <sub>c</sub>
meriympäristön suojele	2	2	N <sub>c</sub> N
miehistö	1	1	N
MIF	1	1	Abb
navigointi	1	1	N
näkyvyys	1	1	N
näkyvyysalue	1	1	N <sub>c</sub>
näkyvyysolosuhde	2	1	N <sub>c</sub>
Off-Shore VTS	2	2	en
off-shore VTS toiminta	1	3	en Abb N
ohjailla	1	1	V
ohjaus	4	1	N
ohjausjärjestelmä	4	1	N <sub>c</sub>
ohjauspalvelu	1	1	N <sub>c</sub>
ohjeistus	2	1	N
onnettomuus	5	1	N
operaattori	9	1	N

pelastus	1	1	N
pelastusviranomainen	0	1	N <sub>c</sub>
päällikkö	2	1	N
päällystö	0	1	N
radio	1	1	N
rahdinkäsittely	1	1	N <sub>c</sub>
rahti	1	1	N
rannikkovaltio	1	1	N <sub>c</sub>
rannikkovartiosto	1	1	N <sub>c</sub>
rantavaltio	1	1	N <sub>c</sub>
raportointimuoto	0	1	N N <sub>c</sub>
Regional Traffic Information Service	2	4	en
RTIS	1	1	Abb
SAR	1	1	Abb
satama	13	1	N
satama-alue	3	1	N <sub>c</sub>
sataman johto- ja informaatiojärjestelmä	1	4	N N C N <sub>c</sub>
sataman käyttöaste	1	2	N N <sub>c</sub>
satamaorganisaatio	1	1	N <sub>c</sub>
satamatoiminto	1	1	N <sub>c</sub>
satamaviranomainen	1	1	N <sub>c</sub>
satamayhteisö	1	1	N <sub>c</sub>
sensori	0	1	N
seuraaminen	2	1	N
sijainninilmaisun	1	1	N <sub>c</sub>
sijainti	1	1	N
sisääntulo	1	1	N <sub>c</sub>
sisääntuloväylä	6	1	N <sub>c</sub>
strateginen liikennekuva	1	2	A N <sub>c</sub>
tehokkuus	9	1	N
tietojen kulku	1	2	N N
tietojen vaihto	5	2	N N
tietokanta	1	1	N <sub>c</sub>
tukiasema	1	1	N <sub>c</sub>
tunnistaminen	2	1	N
turvallisuus	8	1	N
turvallisuusriski	0	1	N <sub>c</sub>
tutka	1	1	N
tutka-asema	7	1	N <sub>c</sub>
tutka-aseman operaattori	1	2	N <sub>c</sub> N
tutka-asetien verkko	1	2	N <sub>c</sub> N
tutkasensori	2	1	N <sub>c</sub>
tutkaverkko	2	1	N <sub>c</sub>
UAIS	1	1	Abb
UAIS-tukiasema	1	1	AbbN
vaarallinen aine	1	2	A N
varastointi	1	1	N
vertikaalinen integraatio	2	2	A N
vertikaalinen tietojenvaihto	1	2	A N <sub>c</sub>
vesikuljetus	1	1	N <sub>c</sub>
vesiliikenne	1	1	N <sub>c</sub>
Vessel Traffic Management and Information Service	2	6	en
Vessel Traffic Management and Information System	1	6	en
Vessel Traffic Service	2	3	en

VHF radio	1	2	Abb N
viranomainen	4	1	N <sub>c</sub>
VTMIS	39	1	Abb
VTMIS toiminta	1	2	Abb N
VTMIS-järjestelmä	1	1	AbbN
VTMIS-palvelu	1	1	AbbN
VTs	49	1	Abb
VTs hallinto	1	2	Abb N
VTs komitea	3	2	Abb N
VTs käsikirja	1	2	Abb N
VTs menetelmä	1	2	Abb N
VTs operaattori	1	2	Abb N
VTs palvelu	1	2	Abb N
VTs resoluutio	2	2	Abb N
VTs säännös	1	2	Abb N
VTs säännöstö	1	2	Abb N
VTs toiminta	8	2	Abb N
VTs-alue	1	1	AbbN
VTs-ohjeistus	1	1	AbbN
VTs-operaattori	1	1	AbbN
VTs-teknologia	1	1	AbbN
VTs-toiminta	1	1	AbbN
VTs-verkko	3	1	AbbN
väylä	1	1	N
yhdistyneet palvelut	1	2	A N
Yhteiset Palvelut	1	2	A N
yhteistoimintakomitea	3	1	N <sub>c</sub>
yhteistyökomitea	2	1	N <sub>c</sub>
yhteistyötoimikunta	1	1	N <sub>c</sub>
ympäristönsuojeluviranomainen	1	1	N <sub>c</sub>
ympäristöriski	1	1	N <sub>c</sub>
ympäristöviranomainen	0	1	N <sub>c</sub>

## LIITE 5

### Venäläisten asiantuntijoiden valitsemat termit tekstistä *Eskos* (107 termiä)

termi	f	sanaluku	kaava
DGPS	1	1	Abb (en)
GPS	1	1	Abb (en)
VTMIS	5	1	Abb (en)
автоматическая идентификационная система	1	3	A A N
агент	3	1	N
АИС	25	1	Abb
бассейн	16	1	N
бассейновая ЕСКУС	3	2	A Abb
бассейновая информационная система	2	3	A A N
бассейновая система	1	2	A N
бассейновый информационный центр	1	3	A A N
бедствие	2	1	N
безопасность мореплавания	7	2	N N
береговая радиолокационная станция	1	3	A A N
береговая система связи	1	3	A N N
вектор скорости	2	2	N N
внутренние территориальные воды	1	3	A A N
гавань	1	1	N
государство флага	1	2	N N
движение судов	2	2	N N
дислокация судов	3	2	N N
единая система контроля и управления судоходством	4	6	A N N C N N
ЕСКУС	49	1	Abb
зона контроля движения судов	2	4	N N N N
зона разделения движения	1	3	N N N
информационная система по надзору за судоходством	1	6	A N P N P N
капитан порта	1	2	N N
квалификация экипажа	1	2	N N
контроль движения судов	1	3	N N N
контроль за движением судов	1	4	N P N N
контроль и управление судоходством	4	4	N C N N
контроль над морским транспортом	1	4	N P A N
контроль над судоходством	4	3	N P N
МАП	2	1	Abb
маячная служба	2	2	A N
мореплавание	7	1	N
морская администрация	4	2	A N
морская администрация порта	1	3	A N N
морская спутниковая связь	1	3	A A N
морской бассейн	1	2	A N
морской диапазон	1	2	A N
морской район	7	2	A N
морской район A1	5	3	A N S
морской район A2	3	3	A N S
морской транспорт	7	2	A N
морской флот	2	2	A N
навигационная система	1	2	A N
надзор за судоходством	3	3	N P N
несение вахты	1	2	N N

обеспечение безопасного плавания	1	3	N A N
обеспечение безопасности мореплавания	4	3	N N N
оператор связи	1	2	N N
открытое море	1	2	A N
подходный путь	1	2	A N
порт	21	1	N
порт назначения	1	2	N N
порт отправления	1	2	N N
портовик	2	1	N
прибрежные воды	0	2	A N
промышленное рыболовство	1	2	A N
радиолокационный контроль за движением судов	1	5	A N P N N
радиомаяк	1	1	N
радиомаяк DGPS	1	2	N Abb
район A1	18	2	N S
район A2	9	2	N S
район A3	5	2	N S
региональная бассейновая информационная система	2	4	A A A N
система обеспечения безопасности мореплавания	2	4	N N N N
система контроля и управления судоходством	4	5	N N C N N
система ОБЧ-радиосвязи	1	2	N N
система по надзору за судоходством	1	5	N P N P N
система судовых сообщений	2	3	N A N
система управления движением судов	1	4	N N N N
Система управления и информационного обеспечения судо- ходства	1	6	N N C A N N
СКП	2	1	Abb
СКЦ	9	1	Abb
служба капитана порта	1	3	N N N
спасательная операция	1	2	A N
спасательно-координационный подцентр	1	2	A N
спасательно-координационный центр	1	2	A N
спутниковая навигационная система	1	3	A A N
спутниковая система связи	1	3	A N N
СПЦ	7	1	Abb
ССС	15	1	Abb
судно	22	1	N
судовладелец	4	1	N
судовое сообщение	3	2	A N
судовой терминал	1	2	A N
судозаход	1	1	N
судопоток	1	1	N
судоходный путь	3	2	A N
судоходство	17	1	N
СУДС	25	1	Abb
территориальные воды	3	2	A N
техническое состояние судна	1	3	A N N
тип судна	1	2	N N
торговое судоходство	1	2	A N
транспондер	1	1	N
транспондерный контроль за движением судов	1	5	A N P N N
управление движением судов	1	3	N N N
управление и информационное обеспечение судоходства	0	5	N C A N N
участник транспортного процесса	5	3	N A N
флот	5	1	N

центр морской спутниковой связи	1	4	N A A N
ЦМСС	2	1	Abb
экипаж	1	1	N
экономическая зона	3	2	A N